ENHANCING BUSINESS VALUE CREATION VIA SOCIAL MEDIA METRICS EVALUATION: A MACHINE LEARNING AND DATA ANALYTICS APPROACH

Olumide marC Adebayo

Under the Supervision of Dr. Elif Kongar

DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN TECHNOLOGY MANAGEMENT

THE SCHOOL OF ENGINEERING

UNIVERSITY OF BRIDGEPORT

CONNECTICUT

December, 2020

ENHANCING BUSINESS VALUE CREATION VIA SOCIAL MEDIA METRICS EVALUATION: A MACHINE LEARNING AND DATA ANALYTICS APPROACH

Approvals

Committee Members

Name	Signature	Date
Dr. Elif Kongar	Elif Kongar	12/7/2020
Dr. Christian Bach	Christian Bach	12/7/2020
Dr. Ruba Deeb		12/7/2020
Dr. Gazi Duman	Gazi Murat Duman	12/7/2020
Dr. Joseph Sarkis		12/7/2020

Ph.D. in Technology Management Program Director

Dr. Elif Kongar Elif Kongar 12/7/2020

Dean, School of Engineering

Dr. Khaled Elleithy Khaled Elleithy 12/7/2020

Dedicated to Jaina, Jesse & Josie, my beloved children.

ENHANCING BUSINESS VALUE CREATION VIA SOCIAL MEDIA METRICS EVALUATION: A MACHINE LEARNING AND DATA ANALYTICS APPROACH

ABSTRACT

Social Media Marketing (SMM) plays an important role in business growth and expansion. With its growing capability and impact, SMM offers a major challenge for decision makers seeking to quantify the value of emergent SMM channels in their marketing mix. The inherent risks and challenges of big data makes the return valuation more pertinent in SMM. Conventional methods used for measuring return on investment (ROI) of marketing activities do not seamlessly translate to SMM operations due to the active involvement of external participants and significant differences in the cost structure. There is no well-established approach to systematically relate organizational social media activities to various revenue streams hindering efforts to justify these investments. This study analyzes social network characteristics and typology to evaluate business performance. Social network typology is a relatively new and important research topic for business performance quantification and evaluation.

Historically, performance measurements in SMM research have always been viewed in terms of numbers of followers, comments, likes, retweets and such. No organization is formed with the end goal of increasing its likes or followers. The main goal of every organization is to increase shareholder value. The contribution of this paper is multi-fold. It aims to start drawing research from these vanity and/or actionable metrics towards organizational performance metrics measurement. The research also introduces a multi-dimensional model that can be instrumental in evaluating the added value of SMM expenditures at the corporate level.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my committee members, Dr. Deeb, Dr. Bach and Dr. Duman for their wise guidance and invaluable feedback. I will forever be grateful for their support. I also extend my unreserved gratitude to Dr. Sarkis, my external committee member, whose knowledge and expertise greatly shaped this research. His thoroughness and contributions were essential in producing a much better research than I could have imagined.

I would also like to express my endless gratitude to my supervisor, Dr. Elif Kongar. I have been extremely blessed to have a supervisor who cares so much about my work, responds to my questions promptly and sees things that I am otherwise oblivious of. Without your balanced drive and patience, I would not be able to complete this program.

Finally, this journey would not have been possible without the unflinching support of my family. Your level of understanding always amazes me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 RESEARCH PROBLEM AND SCOPE

It can be easily argued that the Internet is one of the most disruptive technologies in history. It serves as the basis for many of the currently evolving disruptive technologies e.g. mobile, social media network (SMN), Internet of Things (IoT), Big Data, Artificial Intelligence (AI) and so much more. From a marketing perspective, it provides the platform for almost all virtual distribution channels [1].

Amongst small-to-mid-sized businesses, about 75% engage in one form of social media marketing or the other. Out of these, about 41% post multiple times daily, 23% daily, 27% weekly, 2% monthly and 4% are inconsistent. 88% use social media to boost their brand, 85% use it to start or encourage dialogue with their customer base while 41% use it to reduce research and hiring cost. A further 58% use it to increase sales and alliances with the clientele. [2, 3]. In a recent global report [4], a staggering 65% do not believe they are able to measure the ROI of the social media activities while only 5% say they strongly believe they could. Social media adoption in the marketing mix for SMEs is on the rise and continues to grow due to the proliferation of a plethora of tools and utilities that makes it even easier to engage online [5, 6]. According to a survey carried out by Headley [7], *measuring ROI* and *tying social activities to business outcomes* are the top two most challenging aspects of SMM as shown in Figure 1. As SMEs incorporate and budget for SMM as part of their long term corporate strategies, its return on investment (ROI) becomes questionable [8]. The impact of SMM expenditures becomes an important question [9]. The ability to answer the ROI question threatens the continued buy-in of management in SMEs, who are already crunched on resources. This research explores what has been done by previous researches in this field, highlight the short-comings therein and propose a novel, multi-layered analytical model to start filling this gap.

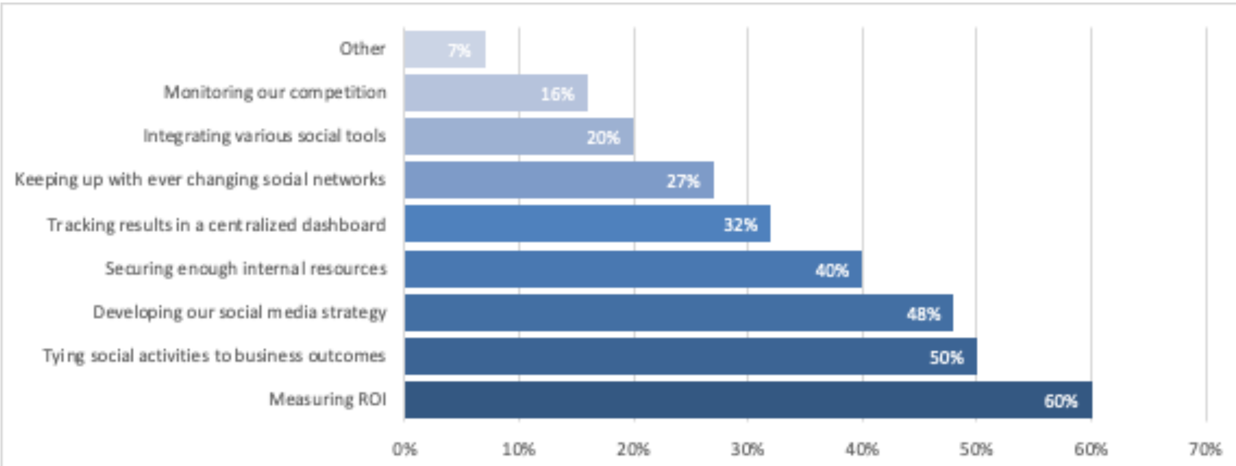


Figure 1. Social Media Marketing Trends

1.2 THE MOTIVATION BEHIND THE RESEARCH

The potential of SMEs to promote domestic-led growth in new and existing industries and to strengthen the resilience of the economy in a competitive and challenging environment is inarguable [10]. According to the World Bank, SMEs represent about 90% of businesses world-wide and account for more than 50% of employment opportunities [11]. The ability of SMEs to appropriately engage in social media and efficiently do so can only be actualized if an instrument that a) measures what is efficient and b) continuously measure and guide SMEs as to how they are deviating from the efficient ways of engaging on social media. This study helps to start filling this gap. As is shown later, the vast majority of research consider *easy pickings* or vain metrics [7, 12, 13]. This research introduces state of the art technology along with proven multi-layered analytical methodologies to determine impactful properties for social media engagement. The motivation for this research comes from the author's love for everything big data, being raised up by a small business owner and interacting daily with siblings who thrive to keep their own SMEs afloat.

1.3 RESEARCH QUESTIONS

The research has attempted to answer the following questions:

- Can we reliably and systematically classify all messaging to fit into a unique class?
    - This question examined the possibility of using machine learning for classifying social media messaging.
- Can we formulate a model that clearly indicates a benefit in some classes above the others?
    - This question investigated the possibility of creating a model that can guide SMEs as to which messaging strategy they should employ in their social media interactions.
- Can this model be transformed into a re-usable, dynamic instrument?
    - This question examined the possibility of implementing this model such that it can adapt to different industries and adapt to changes over time in social media messaging.

1.4 CONTRIBUTION TO THE BODY OF KNOWLEDGE

Starting with the Artificial Intelligence based social media message classification, to building a proven efficiency measurement analysis all the way to a reusable model for determining impactful social property, this study contribute to the literature and practice in the following respects:

- Contributes to the literature on Social Media Marketing performance studies. Unlike previous studies that mostly evaluate performance around *vain metrics*, this study will attempt to re-calibrate Social Media Marketing performance research conversations towards organizational performance as it relates to business performance.

- Contributes to the literature on Small-to-Medium Enterprises by proposing innovative strategies for a) identifying their messaging strategies within their selected social media network, b) evaluating their current efficiency and c) identifying the most impactful messaging strategy on their performance.
- In addition, the discussion of the results helps develop technology management strategies for successful adoption and utilization of Social Media by SMEs. The model builds upon proven efficiency analytical strategy and incorporate time saving state-of-the-art Artificial Intelligence methodologies.

CHAPTER 2 BACKGROUND

2.1 THE SOCIAL MEDIA LANDSCAPE

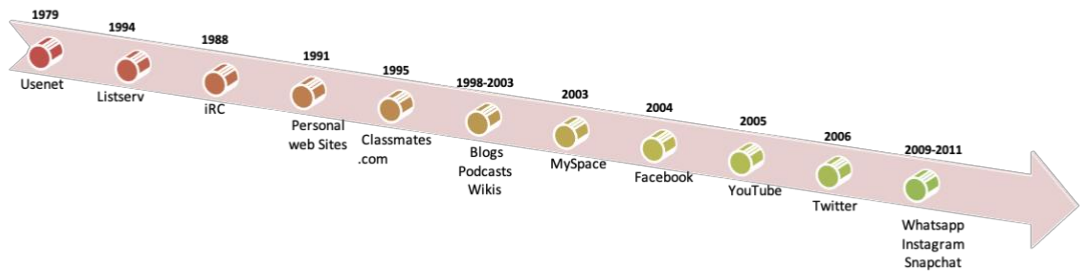


Figure 2. Evolution of Social Networks

The origin of Social Media, as shown in Figure 2 can be traced back to the late 70s with the advent of Usenet [14, 15]. However, the proliferation of mobile devices made possible by increased access to the internet has opened up a totally new era in the social media space. Social Media Networks have become a crucial communication tool for millions of people − in short, they have become ubiquitous [16]. Social Media is defined *as a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content* [14]. It is a place for users and brands to share ideas, provide feedback, state opinions and interests [17, 18]. Social media platforms are unique [19] in that they are *media rich* and empower users to *share* compared to traditional platforms viz; telephone, radio, newspaper and magazines. Users are exposed, at the same time, to plain text, rich text, photos, audio and videos and are able to interact with such content in real-time.

Social media network marketing penetration by US brands has seen a consistent rise since 2013 and currently sits at about 91% for 2019 [20] with a projected US$43 billion social ad spending for the year 2020. As shown in Figure 3**Error! Reference source not found.**, the usage of social media makes it imperative that SMEs engage strongly on social media networks.

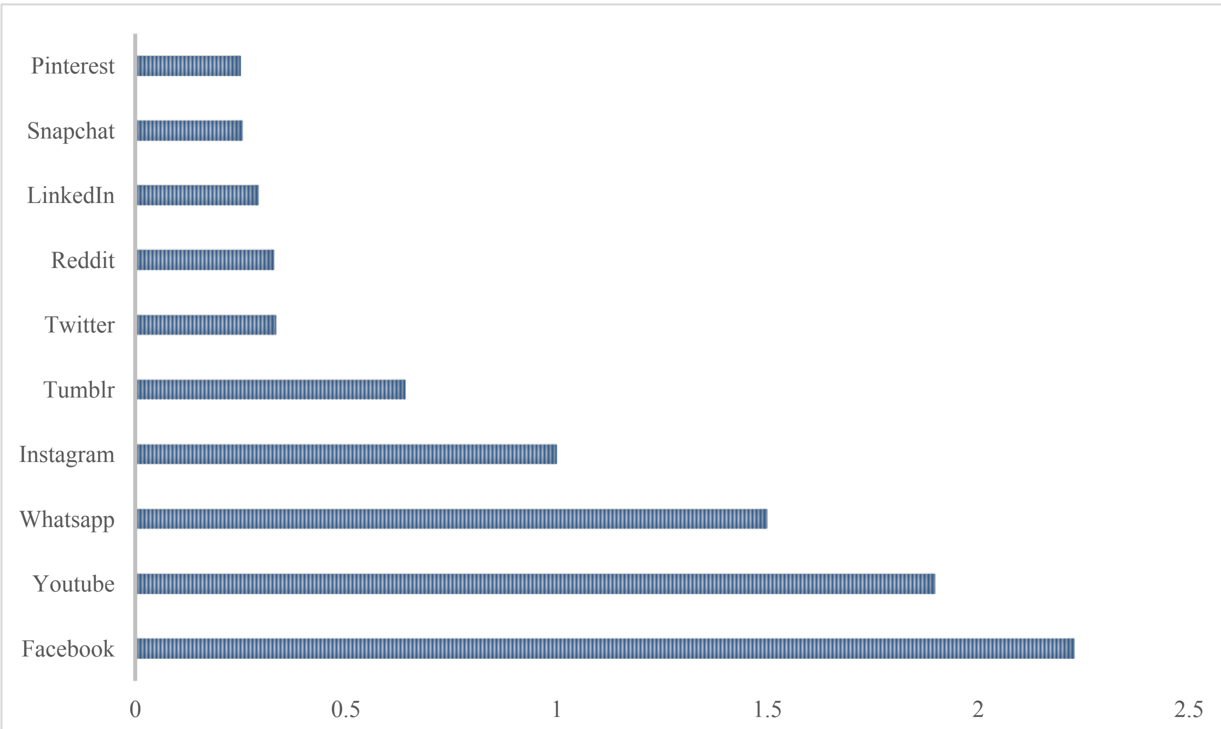


Figure 3. 2019 Monthly Active Users (Billions)

The importance of social media or social networking big data in business intelligence and decision-making cannot be over emphasized. Social media big data utilization has proven itself to be crucial for any business that wishes to survive in today's economy. The utilization involves capturing, analyzing and utilizing collected big data in a timely manner [17]. Today's businesses need to not only value customers but also continuously monitor the impact of social media on their customer base [21].

Brand loyalty is one of the most widely studied SMM research topics. This focus is not surprising since brand loyalty is expected to culminate in brand profitability [22]. It has also been found that user generated social media content has a positive correlation to brand equity and brand attitude variables [9]. These factors are directly related to purchase intention.

User engagement has also seen increasing study [23-26]. User engagement primarily deals with how to motivate, achieve, measure and replicate efficient user participation in social media networks by measuring self-presentation, action, participation, uses, gratifications and experiences of the subject users [27]. This research investigates quantifying or visualizing user engagement due to online marketing strategies [28]. Motivations behind user engagement activities is also important [29, 30]. Significant research on social media big data utilization exists in communications [31], internal operations [32], risk management [33-36], education [35, 37], and User Experience (UX) [37, 38]. But, utilization of social networking sites to evaluate brand objectives is limited [39]. Hoffman and Fodor [8] proposed a bottom-up approach emphasizing the need to consider the motivation of social media utilization. The authors suggested that ROI should include customer social media engagement with the marketer's brand [8].

2.2 SOCIAL MEDIA MARKETING PERFORMANCE EVALUATION

Social Media Marketing needs to be evaluated for its effectiveness in realizing a firm's brand objective [39] thus creating a challenge for marketing managers to document how individual marketing activities could be comprehensively measured in terms of their contribution to the company's financial performance [40]. Milichovsky and Simberova [40] went further to stress that "if the return on marketing investment could be documented, the role of marketing would be significantly elevated throughout the organization. Barring such documentation, marketing will continue to be marginalized" [40].

Social media performance is being addressed in the literature, howbeit, performance is viewed from a non-financial viewpoint [39, 41-43]. As managers and researchers pay more attention to social media, Töllinen and Karjaluoto [41] believes there's a real need for research related to marketing communication performance measurement in the context of social media as there is currently no clear understanding of how effective the media is or how to measure it. Messaging strategy is a concern raised by Ashley and Tuten [42]. According to the authors, marketers have no way of knowing if their messaging is achieving important brand outcomes or improving consumer engagement. Basically, it is like shooting in the dark. An attempt was made by Skulme and Praude [43] to propose a social media return evaluation approach but failed to show or clarify if return was financial (increased sales, increased profit, reduced return, etc) or purely social metrics (increased likes, retweets, etc.).

2.3 MACHINE LEARNING

Big data and data mining research include classifying tweets, comments and various messages on social media networks. As examples, Sinnenberg, Buttenheim [44] classified social media health care data as a precursor to standardized reporting guidelines based on social data. Generalizing their findings, Coursaris, Van Osch [45] indicated that social media marketing message categorization was the foundation for achieving high and positive engagement amongst consumers. The rise of big data includes a plethora of technical analytical tools and skills requirements. Benabderrahmane, Mellouli [46] proposed a textual classification model as the basis of matching skills with job needs across multiple job boards. User identification is a critical element of this activity [47, 48]; this characteristic is separate of content classification which forms the foundation of this research.

Text classification using data mining is an important social media-based research topic. Data mining helps organizations discover hidden knowledge in large data sets [49]. Support Vector Machine (SVM), Bayesian Networks (BN) and Decision Trees (DT) are the top three data mining techniques used by researchers in the area of social media text classification [58]. Biometric, Content Analysis, Cyber Crime, Disease Awareness, Geolocating, Quality Improvement, Risk Management, Semantic Analysis and Sentiment Analysis are prevalent research objectives [58]. Text classifiers can be categorized into lexicon-based and machine learning approaches [50]. Lexicon-based text classification is more dominant in sentiment analysis or opinion analysis where the text is mostly classified into two or three buckets [51]. Lexicon-based sentiment analysis approaches use sentiment lexicons for retrieving the polarity of individual words and aggregate these scores in order to determine the text's polarity (or sentiment) [52]. The need for a pre-defined lexicon (or dictionary) makes this type of text classifier unsuitable for social media data mining beyond simple sentiment derivation as they lack context, are devoid of dynamism and degrade poorly within social media space [53]. Anshari, Alas [37] further states that the performance of lexicon-based sentiment analysis still remains below acceptable levels.

The accuracy of classifiers is one of the major driving forces behind the shift from lexicon-based classifiers to machine learning. Human classification accuracy is rated at around 0.84 [54]. In other words, humans, at first parse, will classify 84% of given texts correctly. With this in mind, any algorithm that can achieve a similar or better accuracy is considered acceptable. While there are many classes of machine learning algorithms, this article concerns itself only with supervised learning. Supervised learning algorithms are set of algorithms that rely on training data sets. They commonly employ three sets of data known as training, validation and test data sets to learn the contextual relationships between the data and the target labels, a.k.a., classes. The training set is used to learn and fit parameters for the classifier; the validation set is used to fine tune what was learnt, and the test set is used to assess the performance of the classifier [55]. Given sizable training data, many machine learning algorithms can achieve or exceed acceptable accuracy levels [56].

In the area of machine learning, Deep Neural Network (DNN) is a class of machine learning algorithms that is more prevalent in social media text classification [57, 58]. Out of these DNN classes, Convolutional Neural Network (CNN) is a robust and efficient Natural Language Processing (NLP) model [59-61]. This research utilizes the Google AutoML, which is also based on CNN [62].

2.4 SENTIMENT ANALYSIS

Sentiment Analysis, or Opinion Mining, of social media data is a fundamental interest of many researches with interesting applications [63]. Sentiment Analysis has been used in a wide variety of applications. Some examples include finance [64, 65], Product Development [66, 67], Marketing [68, 69], Recommendation Systems [70, 71].

There are primarily two approaches to sentiment analysis – machine learning or lexicon-based [72, 73]. Most modern research, and this study, follow the machine learning approach. The machine learning approach can be broken down thus: data collection, pre-processing using Natural Language Processing (NLP) techniques, feature extraction, and model training. NLP processing is used for structuring the text, tokenization, word segmentation, lemmatization, stemming, Part of Speech (POS) tagging, parsing, etc. [73]. Once NLP processing is completed, machine learning techniques are applied to assign polarity weights to entities, topics, themes and categories within a sentence or phrase. A sentiment analysis system for text analysis combines natural language processing and machine learning techniques to assign weighted sentiment scores to the entities, topics, themes and categories within a sentence or phrase [74-76].

In their paper, Pak and Paroubek [77] show that microblogging sites like Twitter and Tumblr lend themselves easily to sentiment analysis. The paper shows the viability of running sentiment analysis on Twitter messages by extracting corpus of words that are positive, negative and/or neutral in polarity. Gokulakrishnan, Priyanthan [78] evaluated multiple classifier algorithms for sentiment analysis like Naïve Bayes, RandomForest, Support Vector Machines and Sequential Mining Optimization. While it highlights the weakness of Bayesian classifiers compared to the non-Bayesian classifiers, the authors also noted that data skewness is a major issue across all classifiers in sentiment analysis.

The rise of Social Media Sentiment Analysis is fueled by a rise in NLP libraries such as SentiStrength, NLTK and Stamford CoreNLP [79-82]; spaCy and SyntaxNet [83]. While different libraries are more optimized for different disciplines, [84] highlights the superiority of NLTK on Social Media data sets. In general, though, Sentiment Analysis on micro-blogging sites like Twitter come with a major drawback in terms of the length of text more so than the informal nature of the content. Piao and Whittle [85] proposed running sentiment analysis on an aggregation of texts rather than individual messages. It has not been shown if this produces a consistent sentiment compared to when individual messages are analyzed.

2.5 DATA ENVELOPMENT ANALYSIS

Data Envelopment Analysis (DEA) is an indispensable tool for quantitative researchers across the world. Emrouznejad and Yang [86] provides a comprehensive listing of all journal published articles relating to DEA since its seminal introduction in 1978. The authors carried out a statistical analysis of article properties on approximately 10,000 published articles in this field. The article along with its accompanying database will serve as an invaluable repository for any researcher performing DEA-based research work. The literature is extremely limited on research that combines data envelopment analysis and social media. The few studies we found on this topic are reviewed below.

In studying the effect of social media adoption on operational performance of tourism sites, [87] employed a three-stage DEA methodology. This layered model approach allows for the elimination of external influences when calculating operational efficiency. Their model differs from all existing models used in determining pure technical efficiency (PTE) in that it avoid *rewarding* poor performance that operate in favorable environment and penalizing good performance that otherwise, operate in unfavorable environment and thus, gives managers an unbiased tool to measure operational efficiency. Using a similar three-stage DEA approach, Martínez-Núñez and Pérez-Aguiar [88] were able to conclude that Online Social Networks (OSN) is an asset that brings efficiency improvements and can be an early indicator of competitive advantage. Their paper introduced a model for assessing the performance of Spanish telecommunication firms who have presence and carry out activities on OSN. Their model helps in providing indicators for the analysis of OSN as scarce, valuable or heterogenous resource.

To measure the effect of the adoption of technological advances across multiple countries, Shao, Lin [89] applied DEA to construct a total factor production performance metric. Schmidt and Hazır [90] applied a variable return to scale DEA model to capture the complex relationship between total cost of a project and its completion time to generate a robust schedule that takes multiple factors into consideration similar to a DEA-based resource allocation model proposed by Lee, Wang [91] for allocating emissions permit.

In order to suggest improvement changes at an engineering department in a e firm, Paradi, Smith [92] used DEA to examine the efficiency, productivity, and effectiveness of the knowledge workers in the department using both the constant returns to scale and variable return to scale models. To gain insight regarding the role of technical experience and task complexity in the efficiency of software development personnel, Otero, Centeno [93] applied DEA to calculate the efficiencies of personnel at a leading software engineering organization. In addition, Talluri and Sarkis [94] show that DEA can be used in identifying alternative high performing solutions by applying cross efficiency analysis to the work of Shafer and Bradford [95].

DEA method is proven to be effective in the evaluation of the efficiency and performance of DMUs when multiple input and output parameters being considered simultaneously. This overcomes a major limitation of other popular efficiency evaluation methods such as Stochastic Frontier Analysis (SFA) [93]. SFA is a parametric approach that requires the specification of a functional form for the frontier in addition to the assumptions about the distributions of the random error and inefficiency error terms making it very restrictive [96].

Another plausible alternative for DEA is the ordinary least squares regression analysis (RA) which can be used either jointly with DEA or individually to assess the comparative performance of such DMUs. However, RA lends itself as a model of choice when comparing the performance of DMUs using a single input or to secure a single output [97]. According to Otero, Centeno [93], considering multiple performance measures simultaneously provides a more thorough evaluation process that structures the decision problem more accurately for practical purposes. The assessment of performance by RA in the more general case where DMUs have multiple inputs and multiple outputs requires the use of simultaneous equations which are fitted to the input-output data [97]. This eventually adds to the computational time without providing any additional benefit.

Therefore, unlike prior social media marketing and performance related analytical studies, this research is an advancement towards the realization of a feasible framework that allows for a meaningful correlation between online social media marketing activities and a firm’s offline performance measurements. It does this by creating an efficient combination of both proven and state-of-the-art methodologies.

CHAPTER 3 RESEARCH PLAN

3.1 LITERATURE REVIEW PLAN

The literature review of this research consists of four major components: Social Media Marketing landscape, Social Media Marketing Performance Evaluation, Machine Learning and Data Envelopment Analysis.

The social media marketing landscape literature reviews the history and growth of social media; usage metrics and current hot research topics in social media network marketing space. The body of literature on Social Media Marketing Performance Evaluation summarizes the efforts so far in terms of researches on various social media marketing metrics and social media marketing performance measurement attempts. The literature review on machine learning addresses interesting areas in social media big data, natural language processing, text classifications and sentiment analysis. It also introduces the main algorithms employed in this space. Finally, the literature review on data envelopment analysis explores work done on DEA for measuring technical efficiencies in similar decision-making units.

The literature review of this research was implemented in three stages: general literature review, targeted literature review, and continuous literature updates. General literature review was carried out during the early stage of the study to help identify the study problems and directions. Following this, the review then targeted certain aspects of the designated methodology. As shown in the Appendix: the literature review of this research builds evaluation matrices to document and analyze the collected materials by identifying the highlights, gap and limitations, research question rationale, methods, implications for future research and practice, themes that emerge, originality/value, and model/techniques.

3.2 DATA COLLECTION PLAN

This research focuses primarily on medium sized organizations rather than small organizations due to the availability and adequacy of data. Our industry selection as depicted in Figure 4 below was by elimination rather than by selection. We eliminated industries that a) are not well represented in the North Eastern United States e.g., Agriculture, Quarrying and Mining b) will pose bureaucratic hurdles to data access e.g., Health, Finance and c) do not fully embrace social media marketing e.g., utilities. Once the industry was selected, brands under the selected industry were searched based on industrial code from LexisNexis. We eliminated brands that did not have social media presence and or readily available periodic financial data.

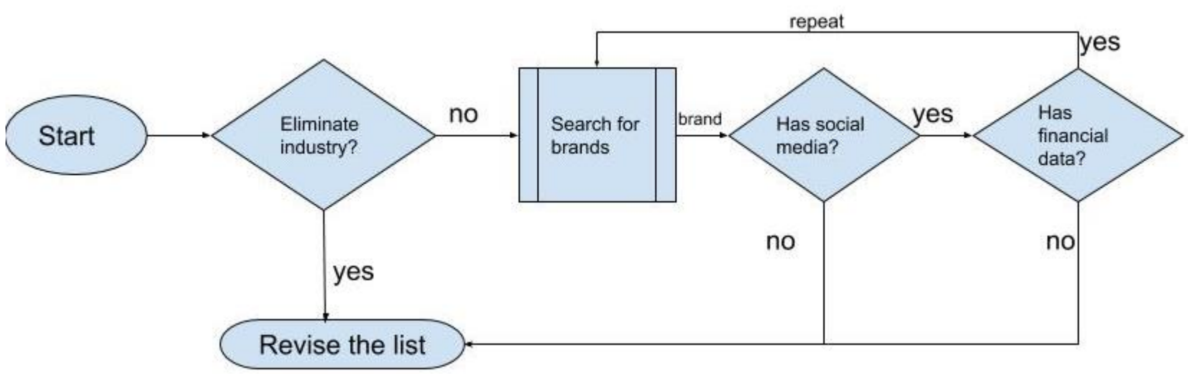


Figure 4. Industry Selection Process

Once the brands have been identified, the actual data collection itself will involve the acquisition of four (4) data sets.

I. Training data set (data set 1): to avoid model bias, this data set will be random data from brands not used in this research. This data set will be used in training our classifier model. It is important that this set represent all classes in an unbiased manner to avoid over-fitting our classifier.

II. Research data set (data set 2): this is the actual data acquired from the selected brands' social media messaging platform.

III. Social metrics data set (data set 3): this is metrics data used in calculating social media efficiency

IV. Organization & Financial metrics data (data set 4): this is data used in calculating business efficiency

3.3 CONCEPTUAL FRAMEWORK WORKING PLAN

The conceptual framework show in Figure 5 defines the participating entities, their actions, and the interactions among them. It addresses the purpose, structure, dependencies, interactions, external conditions, data collection, assumptions, and methodology for the represented system.

The first step is to identify the relevant message typologies, social metrics and organizational metrics relevant to the framework. Once these are identified, the actions and interactions among them will also be identified and described in narrative language, mathematical equations and pseudo-codes in computer language. Based on these interactions, relevant analysis will be evaluated for appropriateness. After the draft high-level framework, the research will be translating the attributes, actions and interactions of entities into a mathematical framework, which will be further translated into a computer programming model.

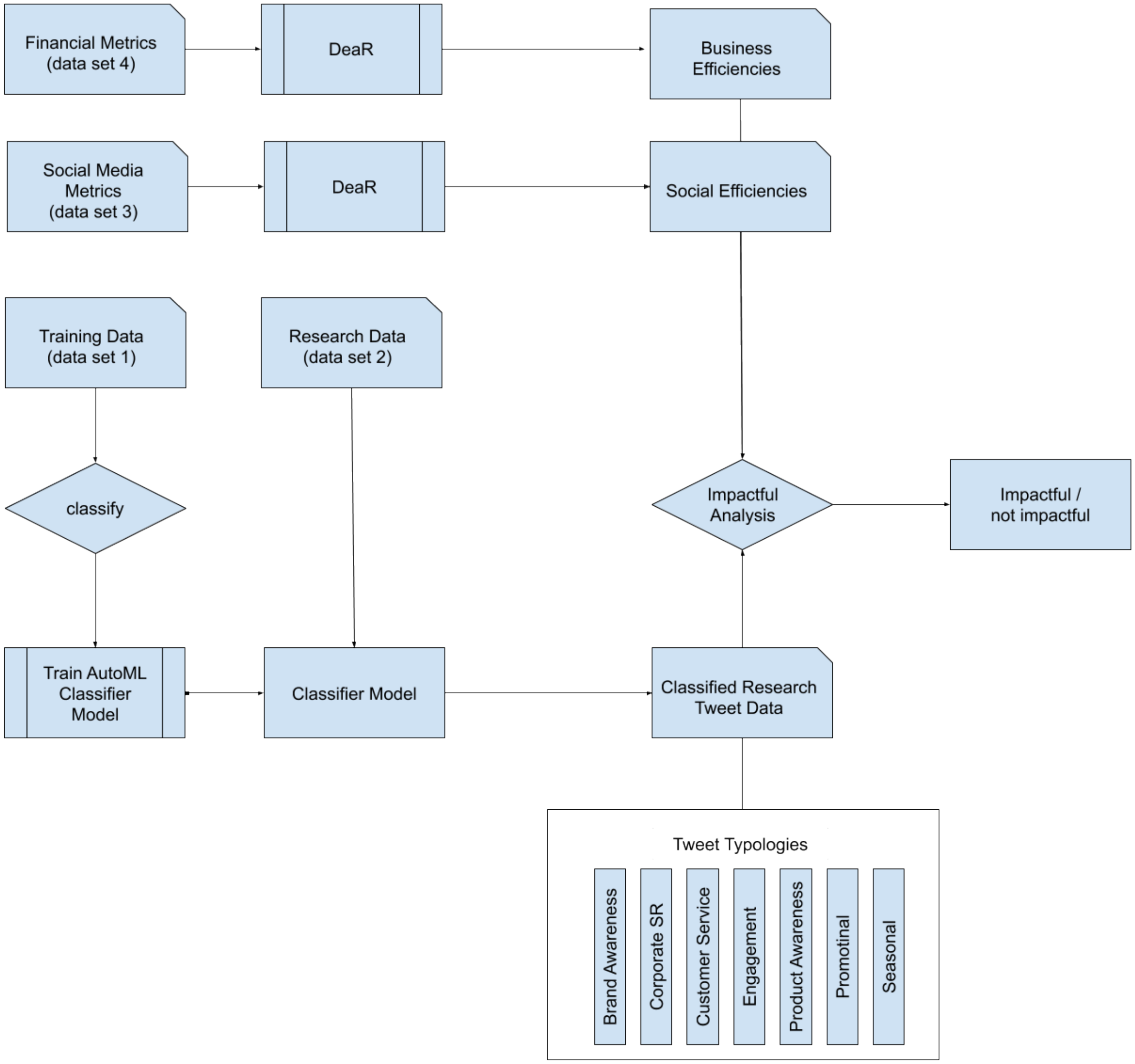


Figure 5. Schematic Representation of the Framework

3.4 MODEL DEVELOPMENT PLAN

Following the constructed conceptual framework, the research will be implemented by developing java-based applications, building cloud-based machine learning models, computing performance analysis using RStudio Version 1.2.1335. As shown in **Error! Reference source not found.**, the p roposed model, starts out by collecting organizational metric data and social metric data.

3.5 PERFORMANCE ANALYSIS PLAN

Following the collection of organizational and financial data, the sample brands will be evaluated for performance using Data Envelopment Analysis (DEA) [98]. DEA models measure the performance of a homogenous set of entities also known as decision making units (DMUs), using multiple input and output variables. DMUs in this study will represent firms. DEA does not assume a parametric relationship or distribution [99]. The performance of each entity is defined by their technical efficiency which can be calculated as the ratio between the weighted outputs and inputs [100]. Generally, input factors are the variables that are subject to minimization, whereas output factors are the ones that are subject to maximization.

The technique is non-parametric because it can provide a quantitative evaluation score for each entity without assuming a parametric relationship between the criteria. DEA uses mathematical programming optimization approach, while statistical methods rely on a maximum likelihood approach with some parametric distribution and relationship amongst the data[99]. The general consensus is that DEA models best perform when there is sufficient number of DMUs in relation to the total number of input and output factors [101-103] and often require the number of DMUs to be at least twice the total number of input and output variables.

Ration-based DEA algorithms can be classified into two categories according to the “orientation” of the model: The first category is called the input-oriented DEA model which concentrate on reducing the amount of input by keeping the output constant. The second category is referred as the output-oriented DEA model which focuses on maximizing the amount of output with the constant amount of input [104-106].

Here, DEA models—as productivity models—may have the following characteristics: Constant Returns to Scale (CRS), or Non-Constant Returns to Scale, i.e., Increasing Returns to Scale (IRS), Decreasing Returns to Scale (DRS), and Variable Returns to Scale (VRS) [107]. Since the focus of this study is on increasing the productivity of each organization an output-oriented CRS DEA model is utilized; for a more balanced approach and sensitivity analysis of productivity solutions—aggressive and benevolent cross-efficiency DEA models are also used [108].

A basic DEA model allows for the introduction of multiple inputs and multiple outputs by obtaining an "efficiency score" of each DMU with conventional output/input weighted productivity ratio analysis. Basic efficiency is the ratio of weighted sum of outputs to the weighted sum of inputs. The relative efficiency score of a test DMU i can be obtained by solving the following DEA ratio model (CCR) proposed by Charnes, Cooper [98].

$E_i = \sum_{k=1}^{K} u_k y_{ki} / \sum_{h=1}^{H} v_h x_{hi}$, where E_i is the efficiency score of DMU i

s.t.

$$\sum_{k=1}^{K} u_k y_{kp} \Bigg/ \sum_{h=1}^{H} v_h x_{hp} \leq 1 \qquad \forall\, p$$

$$u_k, v_h \geq 0 \qquad \forall\; k, h \qquad \text{Eq (1)}$$

where,

i = index of DMU,

y_{ki} = amount of output k for DMU i,

x_{hi} = amount of input h for DMU i,

y_{kp} = amount of output k for DMU p,

x_{hp} = amount of input h for DMU p,

v_h = weight of input h,

u_k = weight of output k.

The linearized version of can be mathematically be expressed as follows:

$$max\ E_i = \sum_{k=1}^{K} u_k y_{ki}$$

$s.t.$

$$\sum_{h=1}^{H} v_h x_{hi} = 1, \qquad \text{Eq (2)}$$

$$\sum_{k=1}^{K} u_k y_{kp} - \sum_{h=1}^{H} v_h x_{hp} \leq 0 \qquad \forall\ p,$$

$$u_k, v_h \geq 0 \qquad \forall\ k, h.$$

Solving Equation (3) provides the relative efficiency score of DMU i relative to the remaining DMUs. A DMU i with an efficiency score of 1 is considered to be efficient whereas any score other than 1 implies that the DMU i is inefficient.

A variant of DEA, DEA with Cross-Efficiency Measurement [109], allows the ranking of DMUs [110] and hence is mostly utilized for peer evaluation; the CCR model usually results in multiple efficient units with a score of 1 and it is difficult to discriminate amongst those DMUs. In addition, the method also eliminates unrealistic weight schemes since it does not require unrealistic weight restrictions [110-112]. Doyle and Green [113] formulated the method as shown in equation (3):

$$\min \sum_{k=1}^{K} u_k \left(\sum_{j=1, j \neq i}^{F} y_{kj} \right) \quad or \quad \max \sum_{k=1}^{K} u_k \left(\sum_{j=1, j \neq i}^{F} y_{kj} \right)$$

s.t.

$$\sum_{h=1}^{H} v_h \left(\sum_{j=1, j \neq i}^{F} x_{hj} \right) = 1,$$

$$\sum_{k=1}^{K} u_k \, y_{ki} - E_i \sum_{h=1}^{H} v_h \, x_{hi} = 0, \qquad \text{Eq(3)}$$

$$\sum_{k=1}^{K} u_k \, y_{ki} - \sum_{h=1}^{H} v_h \, x_{hi} \leq 0, \quad j = 1,2, \ldots., F \; and \; j \neq i \,,$$

$$v_h \,, u_k \geq 0 \,, \quad \forall k, h.$$

When the objective function is to minimize, Eq (3) provides the *aggressive* cross-efficiency DEA formulation and aims at minimizing the sum of weighted outputs. When the objective function is to maximize, the equation is defined as a *benevolent* cross-efficiency formulation and aims at maximizing the sum of the weighted outputs [114]. In addition to the CCR model, this study utilizes the benevolent cross-efficiency formulation to obtain the set of the optimal weights. By applying the cross-efficiency DEA, all the DMUs are assessed by the weights of target DMU *i*. Following this, the average value is calculated. The cross-efficiency score of the j^{th} DMU can be calculated as in Eq (4):

$$CE_j = \frac{\sum_{i=1}^{F} E_{ij}}{F} = \frac{\sum_{i=1}^{F} \left(\dfrac{u_1^i y_{1j} + u_2^i y_{2j} + \cdots + u_K^i y_{Kj}}{v_1^i x_{1j} + v_2^i x_{2j} + \cdots + v_H^i x_{Hj}} \right)}{F} \quad i = 1,2, \ldots, F. \qquad \text{Eq(4)}$$

The ranking of each DMU can then be obtained via its cross-efficiency (CE_j) score since higher values of (CE_j) indicate higher efficiency and vice versa.

3.6 TWEET CLASSIFICATION PLAN

Simply put, classification is the process of separating data into different piles [115, 116]. Before machine learning became mainstream, most researchers either manually classify their data or pay someone to do the classification for them. Amazon's Mechanical Turk was a popular source of cheap labor for data classification and widely used by researchers across the globe [117].

Text classification using machine learning algorithms has started has increasingly become important in social media-based research arena. Data mining helps organizations discover hidden knowledge in large data sets [49]. As mentioned earlier, Support Vector Machine (SVM), Bayesian Networks (BN) and Decision Trees (DT) are the top three data mining techniques used by researchers in the area of social media text classification. Implementing a deep learning text classifier model from scratch comes with a steep learning curve for a non-data scientist. To make ML more accessible to non-data scientists, cloud-based companies have been promoting Automated Machine Learning (autoML) platforms. AutoML attempts to construct machine learning programs without human assistance and within limited computational budgets [118]. [118-121] give a state-of-the-art review of autoML and the inner workings of a cloud-based autoML. For this research, Google Cloud AutoML [122] is utilized due to its compatibility with Google's product landscape, ease of use and lower learning curve. The process of classifying the tweets in the research data set using Google's AutoML is iterative.

When assessing the quality of information retrieval systems, *precision* and *recall* are the two most prevalent performance measures used [123, 124]. Precision is the percentage of relevance. Recall is the percentage of correctness. Even though achieving a high precision and an equally high recall is appealing, it is not always achievable in non-trivial (binary, true/false) cases. A mid-range threshold or confidence level of 0.5 will be selected in order to achieve a balanced precision and recall. A higher confidence level of say, 0.8, will result in a much higher precision but lower recall unless the model has been trained with a far much higher amount of training data. The Accuracy of the classifier can be calculated in terms of its precision and recall as the F_1 score [125]:

$$F1 = \frac{2}{\left(\left(\frac{1}{precision}\right)+\left(\frac{1}{recall}\right)\right)} = \frac{2(precision*recall)}{precision+recall} \qquad Eq(5)$$

The value of equation 5 is accuracy of the classifier and can be improved further by using a larger training data set.

CHAPTER 4 TWEET CLASSIFICATION

4.1 DATA COLLECTION

This process of elimination resulted in two industries for possible research, construction and retail trade. The subsections spread under retail trade helped sway the decision towards its selection. While this study opted for furniture and home furnishings stores, some of the other interesting subsections under this industrial sector include motor vehicle and parts dealers, building material and garden equipment and supplies dealers, food and beverage stores and health and personal care stores, just to list a few. The industry lends itself strongly to medium sized organization status firms and any of the listed sub-sectors could have been easily picked for a reasonable research study.

Brands were systematically selected via two searches in the Nexis Uni®[1] database maintained by LexisNexis. Both searches were limited to US-based firms. The first search was for firms with the identified North American Industry Classification System (NAICS[2]) Code 442110 (furniture stores) and the second search was for firms with the identified SIC[3] Code 5712 (furniture store). 2849 firms overlapped in both sets with 3148 unique firms across both sets. The 3148 were further filtered by dropping records without revenue information and records where either the SIC code or the NAICS was not designated as primary, bringing the number of potential brands down to 192. Out of these 192, 90 firms had Twitter accounts while only 70 have tweeted in 2019. The

[1] LexisUni Database maintained by LexisNexis: https://www.lexisnexis.com/en-us/products/nexis-uni.page

[2] North American Industry Classification System: https://www.naics.com/

[3] Standard Industrial Classification: https://www.osha.gov/pls/imis/sicsearch.html

organizations that provided complete financial data were then selected leaving 43 companies in the dataset. Out of the 43, 20 firms were identified as SMEs, with the remaining 23 being classified as large-sized enterprises (LGEs).

This study utilizes Twitter as its social media network data source. Twitter is recognized as the most popular source for social big data research both in academia and industry [126]. Today, even though Facebook still ranks #1 amongst social media sites due to its extensive utilization for sales, marketing and customer service by many organizations, it does not lend itself readily to educational research. Twitter, however, is considered to be research-friendly with its Application Programming Interfaces (APIs) that make harvesting larger data sets more efficient [127].

The first step in the data collection involved authentication and authorization. Most social media sites require OAuth (https://oauth.net/) or the more recent version OAuth2.0 (https://oauth.net/2/) for secured handshake between the site and the client application requesting data from the site. This step is normally carried out by logging into the developer section of the Twitter website and following the basic steps to retrieve an OAuth access token.

The next step includes gateway selection and structural analysis. The decision to use a gateway or prepackaged software development kit was born out of the need to shorten the development time and also to bypass the Twitter application approval process. For this research, Twurl, an application widely used by client applications to communicate with Twitter [128], was utilized. This was set-up on a Linux Ubuntu 16.04.2 LTS and configured with the OAuth tokens secured in the earlier step.

A custom java application was developed for the Tweet Extraction Process to control Twurl and store the tweets into a data store.

```
Extract Tweets
for each Brand(handle, last id) do{
      ResponseObject = exec Twurl (handl, id)
      if (ResponseObject != null ) then
            TweetObjectArray [] = parse ResponseObject
            for each ( TweetObject in Tweet ObjectArray) do {
                  extract required data
                  persist in database
                  discard object
            }
      end if
}
```

Figure 6. Pseudo Code for the Tweet Extraction Process

Following this initial one-time setup, the actual data acquisition process, shown in Figure 6, involved collecting two sets of data, *i.e.*, training data and research data. For the training data set, 1,000 tweets are collected across randomly selected brands with 700 of them utilized for the actual training. The brand selection carried no weight in the training data set. The only consideration given was to ensure that the brands covered a wide range of sub-industries under the Retail Industry division to afford the model the ability to create a meaningful corpus of text for classifier model training. The second set of data included the 9,700 tweets collected across 70 brands initially selected for this study.

4.2 CLASSIFICATION

There are seven broad categories for classifying SMM messages [45, 129]. These categories include Brand Awareness, Corporate Social Responsibility, Customer Service, [User] Engagement, Product Awareness, Promotional, and Seasonal as shown in Table 1.

Table 1. Definitions of the Tweet Categories

Category	Definition
Brand Awareness	Tweets that provide information regarding the organizational activities such as hires, location openings, celebrity visits, etc.
Customer Service	Tweets that attempt to resolve customer issues.
Corporate Social Responsibility	Tweets that mention charity, goodwill or social consciousness.
Product Awareness	Tweets that draw attention to specific products, product attributes or instructions.
Promotion	Tweets that mention giveaways, contests, sweepstakes, coupons, and reduced rates.
Seasonal	Tweets that mention periodic events such as holidays.
Engagement	Tweets that do not fit into the above six categories.

Ones classes were decided, an interactive Java-based application was developed to present the coder with the database of information regarding each tweet from the training data set and allows the coder to choose the desired classification for a given tweet.

For this research, Google Cloud AutoML [122] is utilized due to its compatibility with Google's product landscape, ease of use and lower learning curve. Following the QuickStart Guide [130] a project was set up. Once the labeled training data set generated above was uploaded, the classifier training was completed in less than 5 hours. The iterative process of classifying the tweets in the research data set using Google's AutoML is shown as pseudocode in Figure 7.

```
Classify Tweet
for each Unclassified Tweet do{
ResponseObject = call Google AutoML (classification model id, tweet text)
if (ResponseObject != null ) then
extract highest scoring class
persist in database
discard object
}
end if
}
```

Figure 7. Pseudocode of the Tweet Classification Process

The sub-system flow for the tweet acquisition and classification is depicted in Figure 8 below.

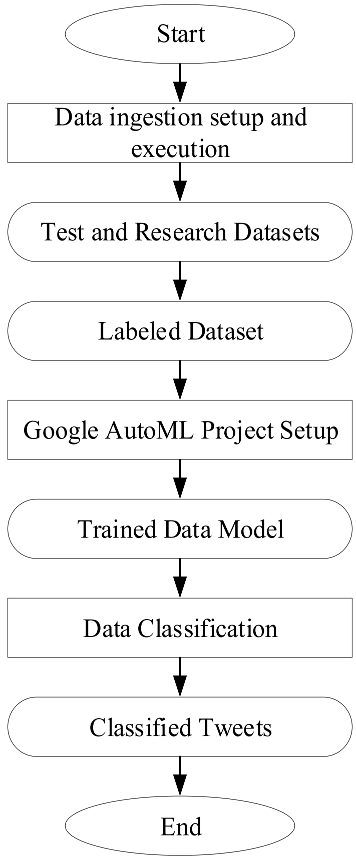


Figure 8. The Tweet Classification Process

Applying Equation 5 to the classifier resulted in an accuracy rate of 82% based on values shown in Table 2. The accuracy can be significantly increased by increasing the size of the training data set.

Table 3 shows the data distribution for the model training, validation and testing data set.

Table 2. Statistical Analysis Results of the Google Cloud AutoML Model

Statistics	Value
Confidence Level	0.5
Training Items	702
Precision	86.9%
Recall	76.9%
Accuracy	82.0%

Table 3. Training Data Distribution

Tweet Category	Training	Validation	Test	Total
Brand Awareness	97	29	30	156
Corporate Social Responsibility	101	27	31	159
Customer Service	101	31	31	163
Engagement	101	31	31	163
Product Awareness	101	31	31	163
Promotion	100	31	31	162
Seasonal	101	24	31	156

After training and deploying the AutoML model for classifying tweets into the research tweets were fed to the model following the algorithm defined in Figure 7. Google AutoML provides a RESTful API for interacting with deployed models. Table 4 shows the collated output of the classifier model. A visual inspection of the output data reveals that not all DMUs participate in all message classes, which is indicated by the cells with 0. Corporate Social Responsibility also seems to be the tweet class with lowest participation. This might be an artifact of the business size or a failure to announce the ‘social good’ that the brand is engaging in. Without applying any statistical or data analytics process, it can be observed that DMUs 21 and above seem to have more tweets and more tweets distributed across the various classes. This is not surprising since these DMUs are the larger brands.

Table 4. Tweet Classification Across All Brands

DMU	Product Awareness	Promotional	Brand Awareness	Seasonal	Engagement	Customer Service	Corporate Social Responsibility
1	96	7	12	26	104	1	0
2	11	4	7	2	21	0	1
3	0	0	1	0	0	0	0
4	5	0	24	3	90	0	0
5	32	70	60	9	79	0	1
6	2	1	0	0	7	0	0
7	0	0	1	1	0	0	0
8	21	6	10	5	40	0	0
9	25	5	12	8	60	1	0
10	2	0	2	0	5	0	0
11	1	0	44	1	17	0	3
12	2	0	10	0	22	0	2
13	0	0	0	0	0	0	1
14	4	8	18	18	160	57	0
15	2	0	4	0	13	0	1
16	16	47	16	14	55	1	2
17	0	0	2	0	0	0	0
18	106	8	39	20	68	8	9
19	30	29	34	3	32	0	0
20	0	1	0	0	0	1	0
21	9	8	33	1	36	12	3
22	19	32	37	23	53	2	1
23	43	21	46	30	96	4	27
24	1	0	1	0	3	0	0
25	27	12	12	6	40	12	4
26	64	17	8	24	53	1	1
27	48	36	29	37	86	4	2
28	22	30	18	29	117	1	2
29	27	34	35	24	80	0	0
30	0	1	14	0	22	247	0
31	39	8	96	23	87	5	6
32	62	130	36	13	42	10	0
33	13	3	13	10	70	75	5
34	23	17	18	24	155	89	15
35	63	10	30	33	105	180	15
36	1	18	7	8	4	0	2
37	2	5	66	4	131	8	4
38	117	45	17	46	110	3	0
39	47	36	19	11	46	56	0
40	1	0	11	2	32	229	0
41	13	5	10	20	105	134	5
42	32	28	1	8	48	260	1
43	13	2	42	7	110	179	1

CHAPTER 5 SENTIMENT ANALYSIS

5.1 SENTIMENT ANALYSIS PROCESS

For Sentiment Analysis purposes, the 1800+ already collected in the earlier process attributable to SMEs will be analyzed for polarity. While there was a need to train a classifier model for tweet classification in CHAPTER 4 due to the nature of the task of classification, Sentiment Analysis has well defined publicly available programing libraries. Some of the most popular libraries include SentiStrength, NLTK and Stamford CoreNLP [79-82]. It is widely agreed that SentiStrength is the most widely adopted sentiment analysis library but NLTK is the most suited for social media data [84].

For this research, a Python3 application was developed on an Ubuntu 18.04.5 LTS. The application processes each of the tweet through TextBlob API [131] as shown in Figure 9. TextBlob is an abstraction API using NLTK and as such is suitable for this research. The python pseudo code is provided below.

```
Polarize Tweet
for each Tweet do{
      _text = extract tweet text
      Sentiment(polarity,context) = TextBlob(_text)
      _polarity = sentiment.polarity
      persist in database
      discard object
}
```

Figure 9. Pseudo Code for Tweet Sentiment Analysis

5.2 SENTIMENT ANALYSIS RESULTS

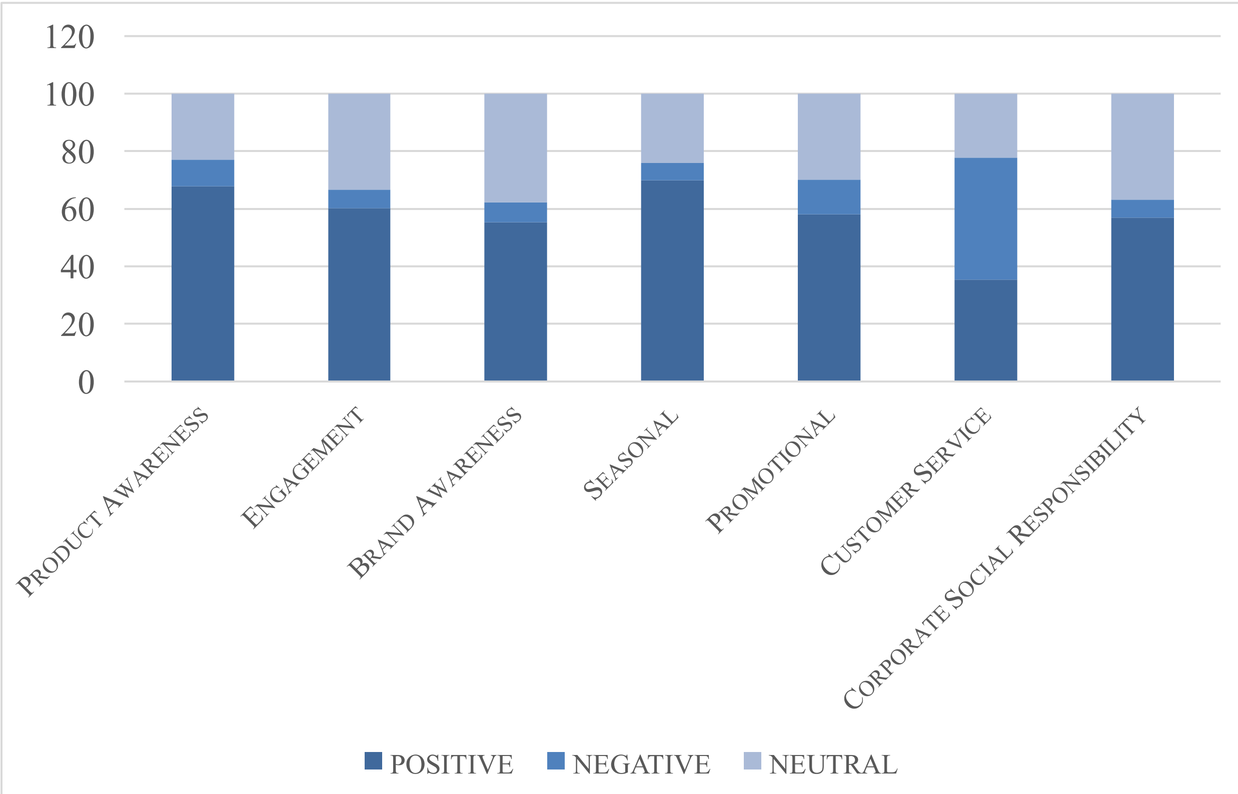


Figure 10. Category Sentiment Analysis for SME

Table 5. Sentiment Analysis Data Table

Category	Positive (%)	Negative (%)	Neutral (%)
Product Awareness	68	9	23
Engagement	60	6	33
Brand Awareness	55	7	38
Seasonal	70	6	24
Promotional	58	12	30
Customer Service	35	42	22
Corporate Social Responsibility	57	6	37

Figure 10 is a visual representation of Table 5. The data in Table 5 is a percentage distribution of sentiments across each class of Twitter message. In other words, under the Product Awareness category, 68% of the tweets where positive, 9% negative and 23% neutral. The data is shown as percentages rather than counts because it allows for a levelled view across all 7 categories. The Customer Service category is an outlier, as expected, due to the nature of the Customer Service category. The category is used mostly for resolving customer issues and as such will have less *gung-ho* type words.

CHAPTER 6 DATA ENVELOPMENT ANALYSIS

6.1 PERFORMANCE ANALYSIS

The R package ‘deaR’ version 1.1.0 is utilized for all calculations using RStudio Version 1.2.1335 by RStudio, Inc. We use six DEA cross-efficiency measurement models to rank furniture stores. These six models together constitute an incremental multistage DEA model [88] providing a detailed analysis of business and SMM performance. The quantitative input and output data for the DEA models are obtained from Twitter [132], the US Manufacturer Survey [133], and D&B Hoovers [134].

The SMM evaluation includes data such as the number of tweets, fans and followers. Corporate success is measured using organizational and financial data. The selection of input and output variables are based on whether the factor was a response or an operational measure. For instance, the number of tweets is considered an input variable since tweets are initiated by the organization, whereas the number of likes derive from customers and hence is classified as an output variable [88]. The complete list of input and output variables utilized in this study are summarized in Table 6Table 6. Input and Output Variables for DEA Models. Each of the factors have been utilized by other studies and the references are provided. A unique factor in this study is list count. List count indicates the number of lists a user belongs to and is included in the model since it is a reliable indicator of engaged Twitter users.

Table 6. Input and Output Variables for DEA Models

DEA Model	**Reference**	**Business Efficiency**	**Social Media Marketing (SMM) Efficiency**
Inputs			
Number of employees	[88, 135-137]	√	√
Total assets	[88, 135, 138, 139]	√	√
Tweets	[88, 135]		√
Outputs			
Annual sales	[88]	√	√
Likes	[135, 140]		√
Followers	[87, 88, 141]		√
Friends	[88]		√
List count			√

Table 7 lists the descriptive statistics of the 43 furniture manufacturers used in the study. With an attempt to avoid misleading weights [142], large differences in dataset is eliminated using the ratio scaling procedure dividing the annual sales and total assets of all organizations by 10,000.

Table 7. Descriptive Statistics of the Companies Used For The DEA Models

	Average	Median	Std. Dev.	Range	Min.	Max.	Total
			SME Business				
# Employees	100.90	90.00	57.00	215.00	10.00	225.00	2018.00
Total Assets	1218.12	700.00	1409.99	5773.40	26.60	5800.00	24362.40
Annual Sales	33105.00	27000.00	27262.57	86700.00	300.00	87000.00	662100.00
			SME Social network				
Tweets	2292.65	1811.00	2147.15	7352.00	17.00	7369.00	45853.00
Likes	479.00	95.50	910.31	3899.00	0.00	3899.00	9580.00
Followers	1368.05	553.00	2319.30	9604.00	6.00	9610.00	27361.00
Friends	676.75	450.00	731.09	2942.00	0.00	2942.00	13535.00
List Count	34.70	11.50	71.74	327.00	0.00	327.00	694.00
			LGE Business				
# Employees	1877.78	875.00	2462.73	10275.00	277.00	10552.00	43189.00
Total Assets	27983.59	9300.00	44206.49	166584.00	6.80	166590.80	643622.57
Annual Sales	447460.87	240000.00	427816.73	1492400.00	7600.00	1500000.00	10291600.00
			LGE Social network				
# Tweets	9366.13	6378.00	10952.55	48222.00	359.00	48581.00	215421.00
# Likes	1947.22	1049.00	2040.95	6555.00	20.00	6575.00	44786.00
# Followers	15666.70	6795.00	26570.37	120299.00	107.00	120406.00	360334.00
# Friends	1379.30	367.00	3207.57	15531.00	22.00	15553.00	31724.00
# List Count	117.04	53.00	190.43	873.00	0.00	873.00	2692.00
			SME+LGE Business				
# Employees	1051.33	300.00	1995.64	10542.00	10.00	10552.00	45207.00
Total Assets	15534.53	2100.00	34741.97	166584.00	6.80	166590.80	667984.97
Annual Sales	254737.21	82000.00	374083.37	1499700.00	300.00	1500000.00	10953700.00
			SME+LGE Social network				
Tweets	6076.14	2478.00	8812.77	48564.00	17.00	48581.00	261274.00
Likes	1264.33	305.00	1762.35	6575.00	0.00	6575.00	54366.00
Followers	9016.16	1125.00	20598.78	120400.00	6.00	120406.00	387695.00
Friends	1052.53	423.00	2399.32	15553.00	0.00	15553.00	45259.00
List Count	78.74	24.00	151.82	873.00	0.00	873.00	3386.00

There are three sets of incremental multistage DEA models employed in the study. The first pair of DEA models include all companies (SMEs and LGEs). The DEA I model measures business efficiency and considers only overall business data. In DEA I, the number of employees and total assets are considered as input variables whereas the annual sales is the output variable. DEA II model introduces social media related variables and includes the number of tweets as an additional input variable whereas the numbers of likes, followers, friends and list count are added as other output variables.

Table 8. DEA I and DEA II model results for simple technical efficiency (CCR), aggressive cross-efficiency (Agg-x), and benevolent cross-efficiency (Ben-x) formulations for SMEs and LGEs

	Business performance DEA I			Social network DEA II		
DMU	CCR	Agg-X	Ben-x	CCR	Agg-X	Ben-x
1	0.03	0.02	0.02	1.00	0.17	0.22
2	0.31	0.23	0.24	1.00	0.40	0.52
3	0.29	0.22	0.23	1.00	0.55	0.65
4	0.34	0.26	0.27	0.73	0.31	0.37
5	0.50	0.39	0.40	1.00	0.56	0.72
6	0.38	0.24	0.25	1.00	0.29	0.33
7	0.19	0.15	0.16	1.00	0.51	0.70
8	0.16	0.13	0.13	0.54	0.17	0.20
9	0.25	0.20	0.21	0.46	0.21	0.30
10	0.22	0.15	0.16	0.51	0.17	0.21
11	0.68	0.53	0.54	0.88	0.36	0.50
12	0.94	0.62	0.64	0.98	0.34	0.50
13	0.71	0.49	0.49	1.00	0.41	0.47
14	0.82	0.76	0.78	1.00	0.43	0.54
15	0.76	0.55	0.56	0.92	0.39	0.55
16	0.36	0.27	0.28	0.42	0.18	0.25
17	0.22	0.19	0.20	1.00	0.26	0.28
18	0.29	0.24	0.24	0.32	0.15	0.18

Table 8 continued.

19	0.37	0.32	0.33	0.79	0.34	0.47
20	0.24	0.19	0.19	0.28	0.13	0.17
21	0.96	0.65	0.67	0.97	0.32	0.46
22	0.14	0.10	0.10	0.33	0.12	0.15
23	0.19	0.17	0.18	0.97	0.26	0.39
24	1.00	0.63	0.62	1.00	0.57	0.66
25	0.52	0.36	0.37	0.56	0.22	0.32
26	0.23	0.20	0.21	0.56	0.19	0.24
27	0.02	0.02	0.02	0.10	0.03	0.04
28	0.22	0.16	0.17	0.31	0.14	0.19
29	0.20	0.15	0.16	0.25	0.11	0.15
30	0.46	0.38	0.39	1.00	0.41	0.59
31	0.23	0.17	0.17	0.66	0.21	0.32
32	0.15	0.12	0.13	0.78	0.25	0.33
33	0.41	0.34	0.34	1.00	0.59	0.73
34	0.35	0.31	0.32	0.43	0.23	0.32
35	1.00	0.92	0.95	1.00	0.48	0.62
36	0.83	0.54	0.56	1.00	0.51	0.66
37	1.00	0.91	0.93	1.00	0.63	0.81
38	0.33	0.23	0.24	0.64	0.24	0.36
39	0.27	0.20	0.21	0.46	0.20	0.28
40	0.43	0.30	0.31	1.00	0.43	0.65
41	0.28	0.26	0.27	0.40	0.21	0.28
42	0.22	0.19	0.19	0.46	0.17	0.21
43	0.04	0.03	0.03	0.15	0.04	0.07

Table 8 provides the DEA I and DEA II model results for simple technical efficiency, aggressive cross-efficiency, and benevolent cross-efficiency formulations. The second pair of DEA models—DEA III and DEA IV—focus solely on SME DMUs with the same input and output variables as DEA I and DEA II models (Table 9).

Table 9. DEA III and DEA IV model results for simple technical efficiency (CCR), aggressive cross-efficiency (Agg-x), and benevolent cross-efficiency (Ben-x) formulations for SMEs

	Business performance DEA III			Social network DEA IV		
Handle	CCR	Agg-X	Ben-x	CCR	Agg-X	Ben-x
1	0.04	0.02	0.03	1.00	0.10	0.21
2	0.37	0.26	0.28	1.00	0.31	0.72
3	0.41	0.29	0.29	1.00	0.43	0.76
4	0.48	0.34	0.34	0.78	0.28	0.46
5	0.71	0.50	0.51	1.00	0.42	0.88
6	0.90	0.38	0.35	1.00	0.40	0.54
7	0.23	0.18	0.19	1.00	0.35	0.76
8	0.22	0.16	0.16	0.59	0.16	0.28
9	0.30	0.24	0.25	0.66	0.21	0.45
10	0.25	0.17	0.19	0.96	0.27	0.36
11	0.82	0.62	0.66	1.00	0.36	0.74
12	1.00	0.65	0.70	1.00	0.37	0.64
13	1.00	0.68	0.66	1.00	0.56	0.87
14	1.00	0.93	0.97	1.00	0.45	0.86
15	0.89	0.62	0.67	1.00	0.49	0.84
16	0.43	0.31	0.33	0.54	0.23	0.41
17	0.26	0.23	0.24	1.00	0.41	0.47
18	0.41	0.30	0.31	0.48	0.20	0.34
19	0.45	0.38	0.40	1.00	0.42	0.75
20	0.29	0.22	0.24	0.40	0.17	0.30

DEA V and DEA VI models constitute the third pair of models and consider only the large sized enterprises (Table 10).

Table 10: DEA V and DEA VI model results for simple technical efficiency (CCR), aggressive cross-efficiency (Agg-x), and benevolent cross-efficiency (Ben-x) formulations for LGEs

	Business performance DEA V			Social network DEA VI		
Handle	CCR	Agg-X	Ben-x	CCR	Agg-X	Ben-x
21	0.96	0.65	0.69	1.00	0.26	0.44
22	0.14	0.10	0.10	0.53	0.19	0.20
23	0.19	0.17	0.18	1.00	0.35	0.50
24	1.00	0.63	0.61	1.00	0.63	0.80
25	0.52	0.36	0.38	0.56	0.17	0.30
26	0.23	0.20	0.21	0.56	0.16	0.21
27	0.02	0.02	0.02	0.14	0.06	0.08
28	0.22	0.16	0.17	0.35	0.14	0.23
29	0.20	0.15	0.16	0.48	0.15	0.23
30	0.46	0.37	0.39	1.00	0.54	0.81
31	0.23	0.17	0.17	0.95	0.31	0.51
32	0.15	0.12	0.12	1.00	0.44	0.59
33	0.41	0.33	0.34	1.00	0.78	0.93
34	0.35	0.30	0.32	0.49	0.25	0.38
35	1.00	0.91	0.95	1.00	0.37	0.54
36	0.83	0.54	0.57	1.00	0.37	0.58
37	1.00	0.89	0.93	1.00	0.61	0.89
38	0.33	0.23	0.24	1.00	0.35	0.63
39	0.27	0.20	0.22	0.46	0.17	0.27
40	0.43	0.30	0.32	1.00	0.43	0.79
41	0.28	0.26	0.27	0.41	0.21	0.31
42	0.22	0.19	0.19	0.46	0.14	0.19
43	0.04	0.03	0.03	0.15	0.04	0.08

As seen in Table 8, Table 9 and Table 10, the CCR results show a large number of efficient DMU with a score of 1; this characteristic makes it difficult to discriminate amongst DMUs—and why we use the cross-efficiency scores. The aggressive cross-efficiency DEA models aim at obtaining a maximum simple efficiency score for a particular DMU while determining a set of weights that will minimize the aggregate output of other DMUs. Alternatively, benevolent DEA models aim at maximizing the aggregate output and is more appropriate for this study. One additional benefit of using the benevolent formulation arises from the fact that the aggressive formulation may cause too much input and output information loss in the efficiency assessment [143]. Therefore, this study considers the results of the benevolent formulation to be more accurate and reliable. Descriptive summary statistics of the six DEA models are provided in Table 11.

Table 11: Summary Results of the DEA Models

Perspective	Business			SMM		
DEA Formulation	CCR	Agg	Ben	CCR	Agg	Ben
SMEs+LGEs	DEA I			DEA II		
Average Efficiency	0.41	0.31	0.32	0.72	0.30	0.39
Standard Deviation	0.29	0.22	0.23	0.30	0.16	0.20
# Efficient DMUs	3	0	0	16	0	0
% Efficient DMUs	6.98%	0.00%	0.00%	37.21%	0.00%	0.00%
Max. Efficiency	1.00	0.92	0.95	1.00	0.63	0.81
Min. Efficiency	0.02	0.02	0.02	0.10	0.03	0.04
SMEs	DEA III			DEA IV		
Average Efficiency	0.52	0.37	0.39	0.87	0.33	0.58
Standard Deviation	0.31	0.22	0.23	0.21	0.12	0.22
# Efficient DMUs	3	0	0	13	0	0
% Efficient DMUs	15.00%	0.00%	0.00%	65.00%	0.00%	0.00%
Max. Efficiency	1.00	0.93	0.97	1.00	0.56	0.88
Min. Efficiency	0.04	0.02	0.03	0.40	0.10	0.21
LGEs	DEA V			DEA VI		
Average Efficiency	0.41	0.32	0.33	0.72	0.31	0.46
Standard Deviation	0.32	0.25	0.26	0.31	0.19	0.26
# Efficient DMUs	3	0	0	11	0	0
% Efficient DMUs	13.04%	0.00%	0.00%	47.83%	0.00%	0.00%
Max. Efficiency	1.00	0.91	0.95	1.00	0.78	0.93
Min. Efficiency	0.02	0.02	0.02	0.14	0.04	0.08

6.2 PERFORMANCE ANALYSIS RESULTS

Comparing the business effectiveness of SMEs in their own class—Models DEA III and IV in Table 12 and Figure 11—show the top ten SMEs within their groups. The remaining lower scored five firms that are characterized by a large number of employees with larger assets. The top ten firms have relatively small total assets and number of employees indicating better utilization of business resources. When SMEs are evaluated for SMM effectiveness, the organizations with strong assets but weaker social media presence are replaced in the top ten rankings by those organizations with higher numbers of follower and friend—despite their relatively limited resources.

Evaluation of LGEs amongst their own size of firms—represented by models DEA V and VI in Table 12 and Figure 11—include the top five companies that were in the industry-at-large comparison in the DEA I model results. The remaining five organizations in the top ten for DEA V, have similar financial standing and are primarily characterized by high annual sales. Introducing SMM metrics—DEA VI model—results in companies with high sales being replaced with organizations that have strong SMM presence; indicated by higher number of Tweets and other related measures such as the number of followers.

CHAPTER 7 RESULTS

When we consider the tweet typologies of the top SME firms, it helps to provide a picture of where the top performers are focused. The top four SME performers in terms of business efficiency (DEA I and DEA III) all show a strong focus on engaging users on social media—an *engagement* typology—accounting for well over 50% of their Twitter activities. They also have a very small footprint for *promotional and product awareness* dimensions. Four other SME DMUs show an interest in engagement along with promotional and product awareness dimensions but are not highly ranked. The key difference between them being that the lower ranked DMUs have few employees. This potentially could be a function of the quality of the engagement where the employees are not social media savvy, or a distraction created with a multi-focused social media messaging campaign. All top ten SMEs barely register on the *seasonal* and *corporate social responsibility* dimensions and as such, we can assume those typographies have no direct impact on the business efficiency rankings.

Applying the Sentiment Analysis results to the above tells an even more interesting story. Apart from Customer Engagement, explained earlier, Promotional and Product Awareness have the worst average sentiment (-12% and -9% respectively) while Engagement ranks with the least negative sentiment (-6%). The neutral sentiment is evenly distributed across all message classes and as such can be discounted as having no significant interpretation from one class to another.

These initial findings mean that SMEs, engaging users on social media might be more valuable than inducing them with promotional offerings as long as the firm can support the engagement with enough manpower. This inference cannot be made for LGE data. LGEs likely have the resources and wherewithal to engage on multiple social channels. This may mean that a more holistic SMM typology is needed when considering LGEs.

The results indicate that DEA results are clearly sensitive to the SMM metrics they employ; they do provide some initial exploratory results that can link the multitude of performance measures. Tying rapidly changing social media data—operational data—to longer term financial indicators provides insights than if separately evaluated. These models and this methodology allow for dynamic and timely measurement of SMM performance.

Table 12. Top ten benevolent cross efficiency results of the DEA Models

SMEs+LGEs				SMEs				LGEs			
Business (DEA I)		SMM (DEA II)		Business (DEA III)		SMM (DEA IV)		Business (DEA V)		SMM (DEA VI)	
DMU	Eff.	DMU	Eff.	DMU	Eff.	DMU	Eff.	DMU	Eff.	DMU	Eff.
35	0.95	37	0.81	14	0.97	5	0.88	35	0.95	33	0.93
37	0.93	33	0.73	12	0.70	13	0.87	37	0.93	37	0.89
14	0.78	5	0.72	15	0.67	14	0.86	21	0.69	30	0.81
21	0.67	7	0.70	13	0.66	15	0.84	24	0.61	24	0.80
12	0.64	36	0.66	11	0.66	3	0.76	36	0.57	40	0.79
24	0.62	24	0.66	5	0.51	7	0.76	30	0.39	38	0.63
15	0.56	3	0.65	19	0.40	19	0.75	25	0.38	32	0.59
36	0.56	40	0.65	6	0.35	11	0.74	33	0.34	36	0.58
11	0.54	35	0.62	4	0.34	2	0.72	34	0.32	35	0.54
13	0.49	30	0.59	16	0.33	12	0.64	40	0.32	31	0.51

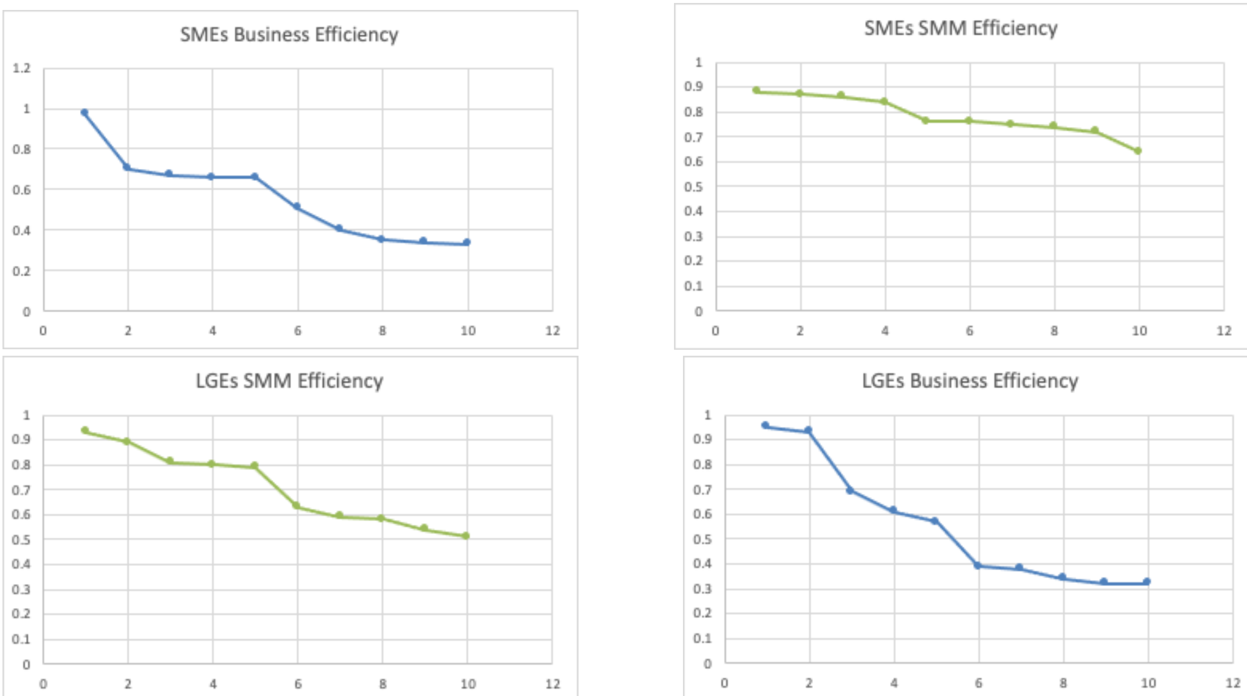


Figure 11. Top ten benevolent cross efficiency results of the DEA Models

CHAPTER 8   CONCLUSIONS

This study presented a conceptual framework and analytical model of selected US-based Furniture Retail Stores for addressing the gap in translating social media marketing and organization performance based on publicly available data from Twitter and Hoover. The proposed innovative framework consists of a typology classification system and a performance evaluation system. The framework also includes a system for evaluating message class sentiments. The proposed framework was then implemented using a combination of R Studio, Java and Python programing languages. The results indicate that DEA results are clearly sensitive to the SMM metrics they employ; they do provide some initial exploratory results that can link the multitude of performance measures. The Sentiment Analysis results gives further credence to the observed *impactful* typology – engagement.

While this is significant, further analysis can be conducted to ascertain the independence of the data classifications using a much wider data set. This will ensure that whatever final conclusions are made for policy or technology management processes take into consideration the dependencies or lack thereof of the various typologies

Finally, the main objective of this study was to provide a systematic tool to organizations with limited marketing resources. The social media and financial data-driven approach provides a versatile and reliable platform for social media performance evaluation and for business improvement.

To the best of our knowledge, this study is unique with its focus on measuring social media performance using business metrics and social media data. In addition to providing the literature with real social media data regarding furniture manufacturers, the study also contributes to the state of the art in multiple ways. The study highlights AutoML as a viable, efficient and effective approach to data analysis (classification) for non-data scientist. This approach will drastically reduce the time researchers spend classifying social media data of any kind. This methodology can be extended to any social media using any set of appropriate typologies.

The study went further and introduced a systematic comparison of relative organizations via a multi-dimensional efficiency analysis which provides efficiency data among most likely competitors in addition to efficiency evaluation compared to the best-in-business companies. Since DEA methodology allows both self- and peer-review evaluation, the findings indicate that DEA was useful in providing insight regarding the changes required in current social media strategies that further drive growth. Other approaches may be used including traditional multi-variate correlative based statistical analysis and potentially Bayesian analyses. Given the data sizes in this study, the use of these other techniques was not methodologically appropriate.

Overall, given the pervasiveness and the growth of SMM, more advanced methodologies are needed for detailed managerial and research investigations. This study adds to the discourse and helps build additional foundation for further investigation.

REFERENCES

1. Prahalad, C.K. and V. Ramaswamy, *Co-opting customer competence.* Harvard business review, 2000. **78**(1): p. 79-90.
2. Rivers, A. *Survey Reveals Social Media Habits of Small Business Owners*. 2017 05/02/2017 [cited 2020 05/20/2020]; Available from: https://www.themarketingscope.com/survey-reveals-social-media-habits-small-business-owners/.
3. Team, M. *The importance of Social Media for SMEs*. [cited 2020; Available from: https://www.adglow.com/blog/the-importance-of-social-media-for-smes.
4. Stelzner, M. *2020 Social Media Marketing Industry Report*. 2020 05/11/2020; Available from: https://www.socialmediaexaminer.com/social-media-marketing-industry-report-2020/.
5. Wamba, S.F. and L. Carter, *Social media tools adoption and use by SMEs: An empirical study*, in *Social media and Networking: Concepts, methodologies, tools, and applications*. 2016, IGI Global. p. 791-806.
6. Jones, P., et al., *Exploring social media adoption in small to medium-sized enterprises in Ireland.* Journal of Small Business and Enterprise Development, 2013.
7. Headley, M., *2015 Social Media Marketing Trends.* 2015.
8. Hoffman, D.L. and M. Fodor, *Can you measure the ROI of your social media marketing?* MIT Sloan Management Review, 2010. **52**(1): p. 41.
9. Schivinski, B. and D. Dabrowski, *The effect of social media communication on consumer perceptions of brands.* Journal of Marketing Communications, 2016. **22**(2): p. 189-214.
10. Aris, N.M., *SMEs: Building blocks for economic growth.* Department of National Statistics, Malaysia, 2007.
11. Bank, W. *Improving SMEs' access to finance and finding innovative solutions to unlock sources of capital.* SMALL AND MEDIUM ENTERPRISES (SMES) FINANCE [cited 2020; Available from: https://www.worldbank.org/en/topic/smefinance.
12. Rogers, R., *Digital traces in context| Otherwise engaged: Social media from vanity metrics to critical analytics.* International Journal of Communication, 2018. **12**: p. 23.
13. Barger, V.A. and L. Labrecque, *An integrated marketing communications perspective on social media metrics.* International Journal of Integrated Marketing Communications, Spring, 2013.
14. Kaplan, A.M. and M. Haenlein, *Users of the world, unite! The challenges and opportunities of Social Media.* Business Horizons, 2010. **53**(1): p. 59-68.
15. Sajithra, K. and R. Patil, *Social media–history and components.* Journal of Business and Management, 2013. **7**(1): p. 69-74.
16. Oh, C., et al., *Beyond likes and tweets: Consumer engagement behavior and movie box office in social media.* Information & Management, 2017. **54**(1): p. 25-37.
17. Mayeh, M., R. Scheepers, and M. Valos. *Understanding the role of social media monitoring in generating external intelligence*. in *ACIS 2012: Location, location, location: Proceedings of the 23rd Australasian Conference on Information Systems 2012*. 2012. ACIS.

18. Kietzmann, J.H., et al., *Social media? Get serious! Understanding the functional building blocks of social media.* Business horizons, 2011. **54**(3): p. 241-251.
19. Du Plessis, T., *Theoretical guidelines for social media marketing communication.* Communicare: Journal for Communication Sciences in Southern Africa, 2010. **29**(1): p. 1-20.
20. Guttmann, A. *Social media marketing usage rate in the United States from 2013 to 2019*. 2019 [cited 2020; Available from: https://www.statista.com/statistics/203513/usage-trands-of-social-media-platforms-in-marketing/.
21. Kim, A.J. and E. Ko, *Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand.* Journal of Business Research, 2012. **65**(10): p. 1480-1486.
22. Erdoğmuş, İ.E. and M. Cicek, *The impact of social media marketing on brand loyalty.* Procedia-Social and Behavioral Sciences, 2012. **58**: p. 1353-1360.
23. Daugherty, T., M.S. Eastin, and L. Bright, *Exploring consumer motivations for creating user-generated content.* Journal of interactive advertising, 2008. **8**(2): p. 16-25.
24. Kunz, W., et al., *Customer engagement in a big data world.* Journal of Services Marketing, 2017. **31**(2): p. 161-171.
25. Muntinga, D.G., M. Moorman, and E.G. Smit, *Introducing COBRAs: Exploring motivations for brand-related social media use.* International Journal of advertising, 2011. **30**(1): p. 13-46.
26. Nov, O., *What motivates wikipedians?* Communications of the ACM, 2007. **50**(11): p. 60-64.
27. Sigerson, L. and C. Cheng, *Scales for measuring user engagement with social network sites: A systematic review of psychometric properties.* Computers in Human Behavior, 2018. **83**: p. 87-105.
28. Oh, J., S. Bellur, and S.S. Sundar, *Clicking, assessing, immersing, and sharing: An empirical model of user engagement with interactive media.* Communication Research, 2018. **45**(5): p. 737-763.
29. Adebayo, O., E. Kongar, and N.J. Sheikh. *Impact of External Stimuli on Social Media Engagement: A SME Perspective*. in *2018 Portland International Conference on Management of Engineering and Technology (PICMET)*. 2018. IEEE.
30. Chi, H.-H., *Interactive digital advertising vs. virtual brand community: Exploratory study of user motivation and social media marketing responses in Taiwan.* Journal of Interactive Advertising, 2011. **12**(1): p. 44-61.
31. Lovejoy, K. and G.D. Saxton, *Information, community, and action: How nonprofit organizations use social media.* Journal of Computer-Mediated Communication, 2012. **17**(3): p. 337-353.
32. Culnan, M.J., P.J. McHugh, and J.I. Zubillaga, *How large US companies can use Twitter and other social media to gain business value.* MIS Quarterly Executive, 2010. **9**(4).
33. Edwards, J.S. and E.R. Taborda, *Using Knowledge Management to Give Context to Analytics and Big Data and Reduce Strategic Risk.* Procedia Computer Science, 2016. **99**: p. 36-49.
34. Guo, S., et al., *A Big-Data-based platform of workers' behavior: Observations from the field.* Accident Analysis & Prevention, 2016. **93**: p. 299-309.
35. Wu, K.-J., et al., *Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties.* Journal of Cleaner Production, 2017. **142**: p. 663-676.

36. Zsidisin, G.A. and M.E. Smith, *Managing supply risk with early supplier involvement: a case study and research propositions.* Journal of supply chain management, 2005. **41**(4): p. 44-57.
37. Anshari, M., Y. Alas, and L.S. Guan, *Developing online learning resources: Big data, social networks, and cloud computing to support pervasive knowledge.* Education and Information Technologies, 2016. **21**(6): p. 1663-1677.
38. He, W., S. Zha, and L. Li, *Social media competitive analysis and text mining: A case study in the pizza industry.* International Journal of Information Management, 2013. **33**(3): p. 464-472.
39. Michaelidou, N., N.T. Siamagka, and G. Christodoulides, *Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands.* Industrial marketing management, 2011. **40**(7): p. 1153-1159.
40. Milichovsky, F. and I. Simberova, *Marketing effectiveness: Metrics for effective strategic marketing.* Engineering economics, 2015. **26**(2): p. 211-219.
41. Töllinen, A. and H. Karjaluoto, *Marketing communication metrics for social media.* International Journal of Technology Marketing, 2011. **6**.
42. Ashley, C. and T. Tuten, *Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement.* Psychology & Marketing, 2015. **32**(1): p. 15-27.
43. Skulme, R. and V. Praude, *Social media evaluation metrics.* Oeconomia Copernicana, 2016. **7**(1): p. 131-142.
44. Sinnenberg, L., et al., *Twitter as a tool for health research: a systematic review.* American journal of public health, 2017. **107**(1): p. e1-e8.
45. Coursaris, C.K., W. Van Osch, and B.A. Balogh. *A Social Media Marketing Typology: Classifying Brand Facebook Page Messages For Strategic Consumer Engagement*. in *ECIS*. 2013.
46. Benabderrahmane, S., et al., *Smart4job: A big data framework for intelligent job offers broadcasting using time series forecasting and semantic classification.* Big Data Research, 2017. **7**: p. 16-30.
47. Giesecke, R., *How do Innovators Network?-The Innovation Capacity of Potentially StrongTies in Individuals' Social Networks.* 2018.
48. Xiong, X., et al. *Social Network User Recommendation Method Based on Dynamic Influence*. in *International Conference on Web Information Systems and Applications*. 2018. Springer.
49. Ngai, E.W., L. Xiu, and D.C. Chau, *Application of data mining techniques in customer relationship management: A literature review and classification.* Expert systems with applications, 2009. **36**(2): p. 2592-2602.
50. Mandal, S. and S. Gupta. *A Lexicon-based text classification model to analyse and predict sentiments from online reviews*. in *2016 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*. 2016. IEEE.
51. Zhang, L., et al., *Combining lexicon-based and learning-based methods for Twitter sentiment analysis.* HP Laboratories, Technical Report HPL-2011, 2011. **89**.
52. Hogenboom, A., et al. *Exploiting emoticons in sentiment analysis*. in *Proceedings of the 28th annual ACM symposium on applied computing*. 2013. ACM.
53. Muhammad, A., N. Wiratunga, and R. Lothian, *Contextual sentiment analysis for social media genres.* Knowledge-Based Systems, 2016. **108**: p. 92-101.

54. Hutto, C.J. and E. Gilbert. *Vader: A parsimonious rule-based model for sentiment analysis of social media text*. in *Eighth international AAAI conference on weblogs and social media*. 2014.
55. Ripley, B.D., *Pattern recognition and neural networks*. 2007: Cambridge university press.
56. Aphinyanaphongs, Y., et al. *Text classification for automatic detection of alcohol use-related tweets: A feasibility study*. in *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014)*. 2014. IEEE.
57. Chen, W.-F. and L.-W. Ku, *Utcnn: a deep learning model of stance classificationon on social media text.* arXiv preprint arXiv:1611.03599, 2016.
58. Ren, Y., et al. *Context-sensitive twitter sentiment classification using neural network*. in *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
59. Kalchbrenner, N., E. Grefenstette, and P. Blunsom, *A convolutional neural network for modelling sentences.* arXiv preprint arXiv:1404.2188, 2014.
60. Kim, Y., *Convolutional neural networks for sentence classification.* arXiv preprint arXiv:1408.5882, 2014.
61. Lai, S., et al. *Recurrent convolutional neural networks for text classification*. in *Twenty-ninth AAAI conference on artificial intelligence*. 2015.
62. Junling, H. Sept 16 2018 [cited 2019; Available from: https://aifrontiers.com/2018/09/16/understand-automl-and-neural-architecture-search/.
63. Alharbi, A.S.M. and E. de Doncker, *Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information.* Cognitive Systems Research, 2019. **54**: p. 50-61.
64. Smailović, J., et al. *Predictive sentiment analysis of tweets: A stock market application*. in *International Workshop on Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*. 2013. Springer.
65. Li, X., et al., *News impact on stock price return via sentiment analysis.* Knowledge-Based Systems, 2014. **69**: p. 14-23.
66. Ireland, R. and A. Liu, *Application of data analytics for product design: Sentiment analysis of online product reviews.* CIRP Journal of Manufacturing Science and Technology, 2018. **23**: p. 128-144.
67. Gräßer, F., et al. *Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning*. in *Proceedings of the 2018 International Conference on Digital Health*. 2018.
68. Rambocas, M. and B.G. Pacheco, *Online sentiment analysis in marketing research: a review.* Journal of Research in Interactive Marketing, 2018.
69. Syed, A.Z. *Applying sentiment and emotion analysis on brand tweets for digital marketing*. in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*. 2015. IEEE.
70. Preethi, G., et al. *Application of deep learning to sentiment analysis for recommender system on cloud*. in *2017 International Conference on Computer, Information and Telecommunication Systems (CITS)*. 2017. IEEE.
71. Sun, L., J. Guo, and Y. Zhu, *Applying uncertainty theory into the restaurant recommender system based on sentiment analysis of online Chinese reviews.* World Wide Web, 2019. **22**(1): p. 83-100.
72. Norsten, T., *Exploring the Potential of Twitter Data and Natural Language Processing Techniques to Understand the Usage of Parks in Stockholm*. 2020.

73. Sun, S., C. Luo, and J. Chen, *A review of natural language processing techniques for opinion mining systems.* Information fusion, 2017. **36**: p. 10-25.
74. Baro, R. and T. Palaoag. *Disaster Sentiment Analysis: Addressing the Challenges of Decision-Makers in Visualizing Netizen Tweets*. in *IOP Conference Series: Materials Science and Engineering*. 2020. IOP Publishing.
75. Saini, R.K. and P. Dangi, *Analysis of the Hotel Reviews using Opinion Mining and Machine Learning Concept.*
76. Mia, M.A.H., *Big data analytics.* 2015.
77. Pak, A. and P. Paroubek. *Twitter as a corpus for sentiment analysis and opinion mining*. in *LREc*. 2010.
78. Gokulakrishnan, B., et al. *Opinion mining and sentiment analysis on a twitter data stream*. in *International Conference on Advances in ICT for Emerging Regions (ICTer2012)*. 2012. IEEE.
79. Thelwall, M., et al., *Sentiment strength detection in short informal text.* Journal of the American society for information science and technology, 2010. **61**(12): p. 2544-2558.
80. Joyce, B. and J. Deng. *Sentiment analysis of tweets for the 2016 US presidential election*. in *2017 IEEE MIT Undergraduate Research Technology Conference (URTC)*. 2017. IEEE.
81. Goel, A., J. Gautam, and S. Kumar. *Real time sentiment analysis of tweets using Naive Bayes*. in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*. 2016. IEEE.
82. Islam, M.R. and M.F. Zibran. *Leveraging automated sentiment analysis in software engineering*. in *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*. 2017. IEEE.
83. Al Omran, F.N.A. and C. Treude. *Choosing an NLP library for analyzing software documentation: a systematic literature review and a series of experiments*. in *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*. 2017. IEEE.
84. Lin, B., et al. *Sentiment analysis for software engineering: How far can we go?* in *Proceedings of the 40th International Conference on Software Engineering*. 2018.
85. Piao, S. and J. Whittle. *A feasibility study on extracting twitter users' interests using nlp tools for serendipitous connections*. in *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*. 2011. IEEE.
86. Emrouznejad, A. and G.-l. Yang, *A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016.* Socio-Economic Planning Sciences, 2018. **61**: p. 4-8.
87. Xu, J., J. Wei, and D. Zhao, *Influence of social media on operational efficiency of national scenic spots in china based on three-stage DEA model.* International Journal of Information Management, 2016. **36**(3): p. 374-388.
88. Martínez-Núñez, M. and W.S. Pérez-Aguiar, *Efficiency analysis of information technology and online social networks management: An integrated DEA-model assessment.* Information & Management, 2014. **51**(6): p. 712-725.
89. Shao, B.B., W.T. Lin, and J.Y. Tsai, *An empirical study of the telecommunications service industries using productivity decomposition.* IEEE Transactions on Engineering Management, 2017. **64**(4): p. 437-449.

90. Schmidt, K.W. and Ö. Hazır, *A Data Envelopment Analysis Method for Finding Robust and Cost-Efficient Schedules in Multimode Projects.* IEEE Transactions on Engineering Management, 2019. **67**(2): p. 414-429.
91. Lee, C.-Y., K. Wang, and W. Sun, *Allocation of emissions permit for China's iron and steel industry in an imperfectly competitive market: a Nash equilibrium DEA method.* IEEE Transactions on Engineering Management, 2019.
92. Paradi, J.C., S. Smith, and C. Schaffnit-Chatterjee, *Knowledge worker performance analysis using DEA: an application to engineering design teams at Bell Canada.* IEEE Transactions on Engineering Management, 2002. **49**(2): p. 161-172.
93. Otero, L.D., et al., *A DEA–tobit analysis to understand the role of experience and task factors in the efficiency of software engineers.* IEEE Transactions on Engineering Management, 2012. **59**(3): p. 391-400.
94. Talluri, S. and J. Sarkis, *Extensions in efficiency measurement of alternate machine component grouping solutions via data envelopment analysis.* IEEE Transactions on Engineering Management, 1997. **44**(3): p. 299-304.
95. Shafer, S.M. and J.W. Bradford, *Efficiency measurement of alternative machine component grouping solutions via data envelopment analysis.* IEEE Transactions on Engineering Management, 1995. **42**(2): p. 159-165.
96. Newhouse, J.P., *Frontier estimation: How useful a tool for health economics?* Journal of health economics, 1994. **13**(3): p. 317-322.
97. Thanassoulis, E., *A comparison of regression analysis and data envelopment analysis as alternative methods for performance assessments.* Journal of the operational research society, 1993. **44**(11): p. 1129-1144.
98. Charnes, A., W.W. Cooper, and E. Rhodes, *Measuring the efficiency of decision making units.* European journal of operational research, 1978. **2**(6): p. 429-444.
99. Bogetoft, P. and L. Otto, *Benchmarking with dea, sfa, and r*. Vol. 157. 2011: Springer Science & Business Media.
100. Sarkis, J. and S. Talluri, *Efficiency measurement of hospitals: issues and extensions.* International Journal of Operations & Production Management, 2002. **22**(3): p. 306-313.
101. Bowlin, W.F., *Measuring Performance: An Introduction to Data Envelopment Analysis (DEA).* Journal of Cost Analysis 1998. **7**: p. 3-27.
102. Dyson, R.G.A., R. ; Camanho, A.S. ; Podinovski, V.V. ; Sarrico, C.S. ; Shale, E.A., *Pitfalls and Protocols in DEA.* European Journal of Operational Research, 2001. **132**: p. 245-259.
103. Golany, B.R., Y. , *An Application Procedure for DEA.* Omega, 1989. **17**: p. 237-250.
104. Kongar, E., Pallis, J.M., and Sobh, T.M, *A Non-parametric Approach for Evaluating the Performance of Engineering Schools.* International Journal of Engineering Education, 2010. **26**(5).
105. Duman, G.M. and E. Kongar, *A data envelopment analysis approach to evaluate the efficiency of service and delivery operations.* American society For engineering education (ASEE), Northeast Section, Northeastern University, Boston, MA, 2015.
106. Duman, G.M., et al., *A holistic approach for performance evaluation using quantitative and qualitative data: A food industry case study.* Expert systems with applications, 2017. **81**: p. 410-422.
107. Banker, R.D., A. Charnes, and W.W. Cooper, *Some models for estimating technical and scale inefficiencies in data envelopment analysis.* Management science, 1984. **30**(9): p. 1078-1092.

108. Wang, Y.-M. and K.-S. Chin, *Some alternative models for DEA cross-efficiency evaluation.* International Journal of Production Economics, 2010. **128**(1): p. 332-338.
109. Sexton, T.R., R.H. Silkman, and A.J. Hogan, *Data envelopment analysis: Critique and extensions.* New Directions for Program Evaluation, 1986. **1986**(32): p. 73-105.
110. Liang, L., et al., *Alternative secondary goals in DEA cross-efficiency evaluation.* International Journal of Production Economics, 2008. **113**(2): p. 1025-1030.
111. Anderson, T.R., K. Hollingsworth, and L. Inman, *The Fixed Weighting Nature of A Cross-Evaluation Model.* Journal of Productivity Analysis, 2002. **17**(3): p. 249-255.
112. Yang, G.-l., et al., *Cross-efficiency aggregation in DEA models using the evidential-reasoning approach.* European Journal of Operational Research, 2013. **231**(2): p. 393-404.
113. Doyle, J. and R. Green, *Efficiency and cross-efficiency in DEA: Derivation, meanings and uses.* Journal of Operational Research Society, 1994. **45**(5): p. 567–578.
114. Falagario, M., et al., *Using a DEA-cross efficiency approach in public procurement tenders.* European Journal of Operational Research, 2012. **218**(2): p. 523-529.
115. Hill, D. and W. Kernochan, *Data Classification.* 2006.
116. Woodbury, C., *The Importance of Data Classification and Ownership.* no. C, 2007: p. 1-4.
117. Narr, S., M. Hulfenhaus, and S. Albayrak, *Language-independent twitter sentiment analysis.* Knowledge discovery and machine learning (KDML), LWA, 2012: p. 12-14.
118. Quanming, Y., et al., *Taking human out of learning applications: A survey on automated machine learning.* arXiv preprint arXiv:1810.13306, 2018.
119. Elshawi, R., M. Maher, and S. Sakr, *Automated Machine Learning: State-of-The-Art and Open Challenges.* arXiv preprint arXiv:1906.02287, 2019.
120. He, X., K. Zhao, and X. Chu, *AutoML: A Survey of the State-of-the-Art.* arXiv preprint arXiv:1908.00709, 2019.
121. Zöller, M.-A. and M.F. Huber, *Survey on Automated Machine Learning.* arXiv preprint arXiv:1904.12054, 2019.
122. *Google Cloud AutoML*. January 2020]; Available from: https://cloud.google.com/automl/.
123. Buckland, M. and F. Gey, *The relationship between recall and precision.* Journal of the American society for information science, 1994. **45**(1): p. 12-19.
124. Raghavan, V., P. Bollmann, and G.S. Jung, *A critical investigation of recall and precision as measures of retrieval system performance.* ACM Transactions on Information Systems (TOIS), 1989. **7**(3): p. 205-229.
125. Flach, P. and M. Kull. *Precision-recall-gain curves: PR analysis done right*. in *Advances in neural information processing systems*. 2015.
126. Ahmed, W. *Using Twitter as a data source: an overview of social media research tools*. 2017; Available from: http://blogs.lse.ac.uk/impactofsocialsciences/2017/05/08/using-twitter-as-a-data-source-an-overview-of-social-media-research-tools-updated-for-2017/.
127. Go, A., R. Bhayani, and L. Huang, *Twitter sentiment classification using distant supervision.* CS224N Project Report, Stanford, 2009. **1**(12).
128. *Using Twurl*. [cited 2019; Available from: https://developer.twitter.com/en/docs/tutorials/using-twurl.
129. Goldkuhl, G. and S. Cronholm, *Adding theoretical grounding to grounded theory: Toward multi-grounded theory.* International journal of qualitative methods, 2010. **9**(2): p. 187-205.

130. *Google Cloud AutoML QuickStart Guide*. [cited 2020 January 2020]; Available from: https://cloud.google.com/natural-language/automl/docs/quickstart?authuser=1.
131. Loria, S., *textblob Documentation.* Release 0.15, 2018. **2**.
132. *Twitter*. [cited 2018; Available from: https://twitter.com/.
133. Census, *Annual Survey of Manufactures: General Statistics: Statistics for Industry Groups and Industries: 2016 and 2015*, in *2016 Annual Survey of Manufactures*.
134. *D&B Hoovers*. [cited 2020; Available from: http://www.hoovers.com/.
135. Hermoso, R., M.P. Latorre, and M. Martinez-Nuñez, *Multivariate Data Envelopment Analysis to Measure Airline Efficiency in European Airspace: A Network-Based Approach.* Applied Sciences, 2019. **9**(24): p. 5312.
136. Rouyendegh, B.D., A. Yildizbasi, and I. Yilmaz, *Evaluation of retail ındustry performance ability through ıntegrated ıntuitionistic fuzzy TOPSIS and data envelopment analysis approach.* Soft Computing: p. 1-12.
137. Arjomandi, A. and J.H. Seufert, *An evaluation of the world's major airlines' technical and environmental performance.* Economic Modelling, 2014. **41**: p. 1-12.
138. Sigala, M., *The information and communication technologies productivity impact on the UK hotel sector.* International journal of operations & production management, 2003: p. 1224-1245.
139. Lee, B.L. and A.C. Worthington, *Technical efficiency of mainstream airlines and low-cost carriers: New evidence using bootstrap data envelopment analysis truncated regression.* Journal of Air Transport Management, 2014. **38**: p. 15-20.
140. Pinheiro, M.D. and A. Grilo, *Assessing business efficiency in the use of social networking sites: a DEA approach.* International Proceedings of Economics Development and Research, 2013. **59**: p. 26.
141. Serrano-Cinca, C., Y. Fuertes-Callén, and C. Mar-Molinero, *Measuring DEA efficiency in Internet companies.* Decision Support Systems, 2005. **38**(4): p. 557-573.
142. Cooper, W.W., et al., *Avoiding large differences in weights in cross-efficiency evaluations: application to the ranking of basketball players.* Journal of CENTRUM Cathedra: The Business and Economics Research Journal, 2011. **4**(2): p. 197-215.
143. Wang, Y. and S. Wang, *Approaches to determining the relative importance weights for cross-efficiency aggregation in data envelopment analysis.* Journal of the Operational Research Society, 2013. **64**(1): p. 60-69.

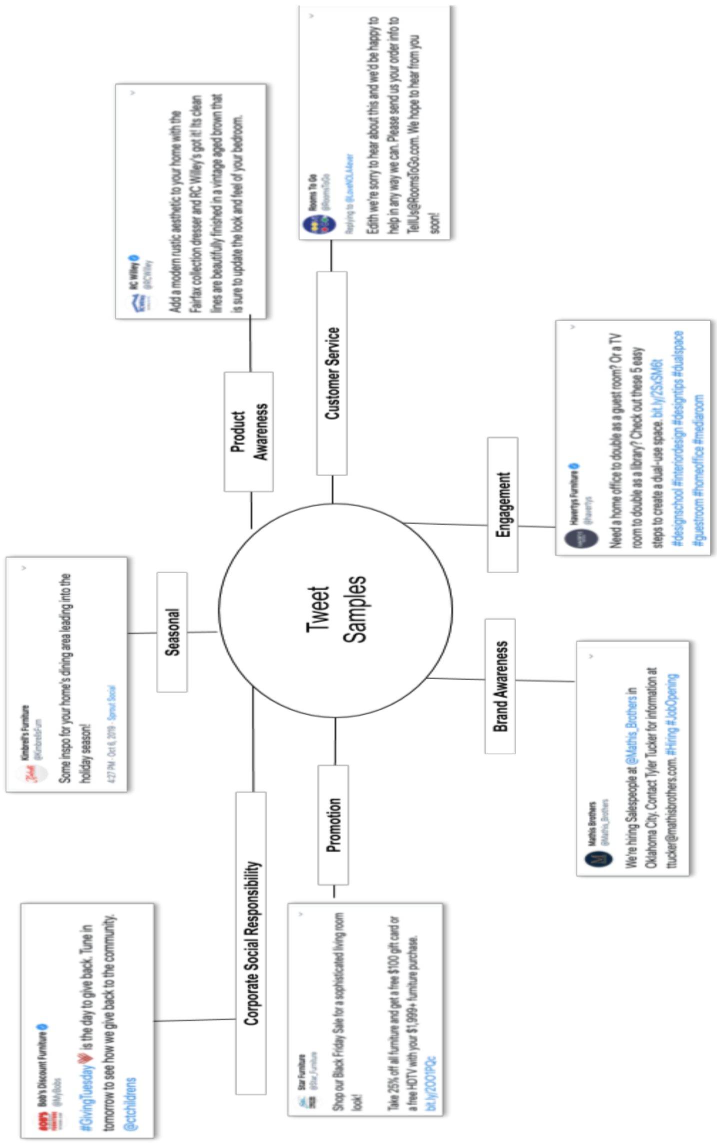
APPENDIX 1. SAMPLE RAW TWEET


```
{

"created_at": "Tue Oct 27 23:06:22 +0000 2020",
 "id": 1321226763322740738,
 "id_str": "1321226763322740738",
 "full_text": "@jbtpba Hi Jonathan,\n\nWe've sent you a DM with the next steps so that our team may address this with you. Please reply in the private thread when you get a moment.\n\n--Megan",
 "truncated": false,
 "entities": {
  "hashtags": [],
  "symbols": [],
  "user_mentions": [
   {
    "screen_name": "jbtpba",
    "name": "Jonathan Tripp",
    "id": 918823577721466881,
    "id_str": "918823577721466881",
    "indices": [0,7]
   }
  ],
 },
 "metadata": {
  "iso_language_code": "en",
  "result_type": "recent"
 },
 "source":"<a href=\"https:\/\/www.spredfast.com\" rel=\"nofollow\">Khoros Marketing<\/a>",
 "in_reply_to_status_id": 1321223062520553477,
 "in_reply_to_status_id_str": "1321223062520553477",
 "in_reply_to_user_id": 918823577721466881,
 "in_reply_to_user_id_str": "918823577721466881",
 "in_reply_to_screen_name": "jbtpba",
 "user": {
  "id": 18455097,
  "id_str": "18455097",
  "name": "Mattress Firm",
  "screen_name": "MattressFirm",
  "location": "",
  "description": "We make it easy to get a great night’s sleep 😴",
  "url": "https:\/\/t.co\/TrSu1iaH23",
  "entities": {
   "url": {
    "urls": [
     {
      "url": "https:\/\/t.co\/TrSu1iaH23",
      "expanded_url": "http:\/\/www.mattressfirm.com",
      "display_url": "mattressfirm.com",
      "indices": [
       0,
       23
      ]
     }
    ]
   },
   "description": {
    "urls": []
   }
  },
```

```
    "protected": false,
    "followers_count": 40860,
    "friends_count": 420,
    "listed_count": 101,
    "created_at": "Mon Dec 29 18:10:57 +0000 2008",
    "favourites_count": 3273,
    "utc_offset": null,
    "time_zone": null,
    "geo_enabled": true,
    "verified": true,
    "statuses_count": 50793,
    "lang": null,
    "contributors_enabled": false,
    "is_translator": false,
    "is_translation_enabled": false,
    "profile_background_color": "FFFFFF",
    "profile_background_image_url": "http:\/\/abs.twimg.com\/images\/themes\/theme10\/bg.gif",
    "profile_background_image_url_https": "https:\/\/abs.twimg.com\/images\/themes\/theme10\/bg.gif",
    "profile_background_tile": false,
    "profile_image_url": "http:\/\/pbs.twimg.com\/profile_images\/964223959708319744\/ZcR0DBk0_normal.jpg",
    "profile_image_url_https":
"https:\/\/pbs.twimg.com\/profile_images\/964223959708319744\/ZcR0DBk0_normal.jpg",
    "profile_banner_url": "https:\/\/pbs.twimg.com\/profile_banners\/18455097\/1517409776",
    "profile_link_color": "0084B4",
    "profile_sidebar_border_color": "FFFFFF",
    "profile_sidebar_fill_color": "DDEEF6",
    "profile_text_color": "333333",
    "profile_use_background_image": true,
    "has_extended_profile": false,
    "default_profile": false,
    "default_profile_image": false,
    "following": true,
    "follow_request_sent": false,
    "notifications": false,
    "translator_type": "none"
  },
  "geo": null,
  "coordinates": null,
  "place": null,
  "contributors": null,
  "is_quote_status": false,
  "retweet_count": 0,
  "favorite_count": 0,
  "favorited": false,
  "retweeted": false,
  "lang": "en"
}
```

APPENDIX 2. SAMPLE TWEET CLASSIFICATION



APPENDIX 3. TWEET COLLECTION JAVA APPLICATION

```
package org.olumide;

import java.io.*;
import java.util.*;
import java.text.*;
import org.apache.log4j.Logger;
import com.google.gson.*;
import java.sql.Connection;
import java.sql.PreparedStatement;


public class Main{
        private static Logger _log = Logger.getLogger(Main.class.getName());

        private static final Gson gson =
                new GsonBuilder().setPrettyPrinting()
                .setFieldNamingPolicy(FieldNamingPolicy.LOWER_CASE_WITH_UNDERSCORES)
                                .create();


        public static final String
twurlCmd="/1.1/statuses/user_timeline.json?screen_name=SCREENER&since_id=MAXID&count=500&include_
rts=1&trim_user=1";


        private static Connection connection = null;

        public static void main(String[] args) throws Exception{

                Runtime rt = Runtime.getRuntime();
                String command =  twurlCmd;

                connection = getDB();

                List<String> handles = getHandles();

                for(String handle: handles){

                        _log.info("PROCESSING "+handle);

                        String _command = command.replace("SCREENER",handle);
                        String _maxId=getMaxId(handle);
                        _command = _command.replace("MAXID",_maxId);

                        _log.info( _command);

                        Process proc = rt.exec(new String[]{"/usr/local/bin/twurl",_command});

                        BufferedReader stdInput =
                                 new BufferedReader(new InputStreamReader(proc.getInputStream()));
                        BufferedReader stdError =
                                new BufferedReader(new InputStreamReader(proc.getErrorStream()));

                        // read the output from the command
                        _log.info("Here is the standard output of the command:\n");
                        String s = null;

                        StringBuffer tweets = new StringBuffer();

                        while ((s = stdInput.readLine()) != null) {
                                //_log.info(s);
                                tweets.append(s);
                        }
```

```
try{

                    JsonArray jarray = parse(tweets.toString());
                    if( jarray != null && jarray.size()>=1){
                            for(int i=0;i<jarray.size();i++){
                                    process(handle,jarray.get(i),i);
                            }
                    }else{
                            continue;
                    }
                    }catch(Exception e){
                            continue;
                    }
                    //try{ }catch{ continue; }

                    // read any errors from the attempted command
                    _log.info("Here is the standard error of the command (if any):\n");
                    while ((s = stdError.readLine()) != null) {
                            _log.error(s);
                    }
            }

    }


    private static void process(String handle,JsonElement je,int idx){
            if( je == null){
                    return;
            }

            JsonObject jobj = je.getAsJsonObject();
            JsonElement _je = jobj.get("id");
            long _id = _je.getAsLong();
            _log.info( handle +" : "+ idx +" : "+ _id);

            String created = jobj.get("created_at").getAsString();
            created = getDate(created);

            String tweet= jobj.get("text").getAsString();
            tweet = clean(tweet);

            int favorite_count = 0;
            int reply_count = 0;
            int quote_count = 0;
            int retweet_count = 0;
            try{ retweet_count= jobj.get("retweet_count").getAsInt(); }catch(Exception e){ }
            try{ quote_count= jobj.get("quote_count").getAsInt(); }catch(Exception e){ }
            try{ reply_count= jobj.get("reply_count").getAsInt(); }catch(Exception e){ }
            try{ favorite_count= jobj.get("favorite_count").getAsInt(); }catch(Exception e){ }

            JsonObject userObj = jobj.get("user").getAsJsonObject();
            String user_id  = userObj.get("id").getAsString();

            String inReply1 = null;
            String inReply2 = null;

            if( !( jobj.get("in_reply_to_status_id").isJsonNull())){
                    inReply1 = jobj.get("in_reply_to_status_id").getAsString();
            }
            if( !( jobj.get("in_reply_to_user_id").isJsonNull())){
                    inReply2 = jobj.get("in_reply_to_user_id").getAsString();
            }
```

```
String in_reply="N";
if( inReply1 != null || inReply2 != null){
        in_reply="Y";
}

int user_mentions= 0;
int media_count = 0;
int hashtags= 0;

JsonObject entities = jobj.get("entities").getAsJsonObject();
if( entities!= null && !entities.isJsonNull() && entities.isJsonObject()){

        if(entities.get("media") != null && ! (entities.get("media").isJsonNull())){

                JsonArray mediaArray = entities.get("media").getAsJsonArray();
                if( mediaArray != null && !mediaArray.isJsonNull()
                && mediaArray.isJsonArray() ){
                        media_count = mediaArray.size();
                }
        }
        if( entities.get("user_mentions") != null
                && ! (entities.get("user_mentions").isJsonNull())){

                JsonArray userMArray =
                        entities.get("user_mentions").getAsJsonArray();
                if( userMArray!= null && !userMArray.isJsonNull()
                        && userMArray.isJsonArray() ){
                        user_mentions = userMArray.size();
                }
        }
        if( entities.get("hashtags") != null
                && ! (entities.get("hashtags").isJsonNull())){
                JsonArray hashArray = entities.get("hashtags").getAsJsonArray();
                if( hashArray!= null && !hashArray.isJsonNull()
                        && hashArray.isJsonArray() ){
                        hashtags = hashArray.size();
                }
        }
}

String sql = "insert into furniture_tweets ";
sql += "(handle,created ,id ,tweet ,media_count ,in_reply ,user_id ";
sql += ",retweet_count ,favorite_count ,reply_count ";
sql += " ,quote_count,user_mentions,hashtags)";
sql += " values ";
sql += "( ?,?, ?,?,?,?,?,?,?,?,?,?,? ) ";

String delim=" _";
StringBuffer _buf = new StringBuffer();
_buf.append(handle);
_buf.append(delim);
_buf.append(created);
_buf.append(delim);
_buf.append(_id);
_buf.append(delim);
_buf.append(tweet);
_buf.append(delim);
_buf.append(media_count);
_buf.append(delim);
_buf.append(in_reply);
_buf.append(delim);
_buf.append(user_id);
_buf.append(delim);
_buf.append(retweet_count);
_buf.append(delim);
_buf.append(favorite_count);
_buf.append(delim);
_buf.append(reply_count);
_buf.append(delim);
_buf.append(quote_count);
```

```
        _buf.append(delim);
        _buf.append(user_mentions);
        _buf.append(delim);
        _buf.append(hashtags);

        PreparedStatement ps = null;
         try {
                 ps = connection.prepareStatement(sql);

                ps.setString(1, handle);
                ps.setString(2, created);
                ps.setLong(3,_id);
                ps.setString(4,tweet);
                ps.setLong(5,media_count);
                ps.setString(6,in_reply);
                ps.setString(7,user_id);
                ps.setLong(8,retweet_count);
                ps.setLong(9,favorite_count);
                ps.setLong(10,reply_count);
                ps.setLong(11,quote_count);
                ps.setLong(12,user_mentions);
                ps.setLong(13,hashtags);

                ps.execute();
         } catch (Throwable t) {
                 t.printStackTrace();
         } finally {
                 try {
                         ps.close();
                 } catch (Throwable t) {
                         t.printStackTrace();
                 }
         }
}

private static JsonArray parse(String s) throws Exception{
        if( isEmpty(s)){
                return null;
        }
        return new JsonParser().parse(s).getAsJsonArray();
}


private static final String TWITTER_DATE="EEE MMM dd HH:mm:ss ZZZZZ yyyy";

private static SimpleDateFormat sf = new SimpleDateFormat(TWITTER_DATE,Locale.ENGLISH);
```

```
private static String getDate(String s){
        try{

        sf.setLenient(true);
        Date date = sf.parse(s);
        String newstring = new SimpleDateFormat("yyyy-MM-dd").format(date);
        return newstring;
        }catch(Exception e){
                e.printStackTrace();
                return s;
        }

}
//remove errant chars
public static String clean(String s){
       if( s == null){
              return s;
       }
        s = s.replaceAll("[\\p{Cc}\\p{Cf}\\p{Co}\\p{Cn}]", "");
        s = s.replaceAll("\\P{InBasic_Latin}", "");
        s = s.replaceAll("\\t","");
        s = s.replaceAll("\\r","");
        s = s.replaceAll("\\n","");
        s = s.replaceAll("\\r\\n","");
        s = s.replaceAll("\\n\\r","");
        return s;
}

private static boolean isEmpty(String str){
        return (str == null || str.trim().length() == 0);
}


private static String getMaxId(String handle){

       String sql = "select coalesce(max(id),1) id
              from furniture_tweets where handle=?";
       String response = null;
       java.sql.ResultSet rs = null;
       PreparedStatement ps = null;
       try {
                ps = connection.prepareStatement(sql);
                ps.setString(1,handle);
              rs = ps.executeQuery();
                if(rs.next()) {
                        response = rs.getString("id");
                }
        } catch (Throwable t) {
                t.printStackTrace();
        } finally {
                try {
                        rs.close();
                        ps.close();
                } catch (Throwable t) {
                        t.printStackTrace();
                }
        }

       _log.info(" max id for "+handle +" is "+response);
       return response;
}

private static void closeDB(Connection conn) throws Exception{
       conn.close();
}
private static Connection getDB() throws Exception{
       String url = "jdbc:mysql://localhost:3306/<DATABASE NAME>";
       String username = "<USERNAME>";
       String password = "<PASSWORD>";
```

```
        try {

                _log.info("loading driver!");
                Class.forName("com.mysql.jdbc.Driver");

                _log.info("Connecting database...");
                Connection connection=java.sql.DriverManager.getConnection(
                        url, username, null) ;

                _log.info("Database connected!");
                return connection;

        } catch (Exception e) {
                throw new IllegalStateException("Cannot connect the database!", e);
        }

    }

    private static List<String> getHandles(){

        List<String> handles = new ArrayList<>();
        //String sql = "select handle from furniture_merchants ";
        String sql = "select TwitterHandle handle from brands ";
        String response = null;
        java.sql.ResultSet rs = null;
        PreparedStatement ps = null;
        try {
                 ps = connection.prepareStatement(sql);
                rs = ps.executeQuery();
                 while(rs.next()) {
                         handles.add(rs.getString("handle"));
                 }
         } catch (Throwable t) {
                 t.printStackTrace();
         } finally {
                 try {
                         rs.close();
                 } catch (Throwable t) {
                         t.printStackTrace();
                 }
                 try {
                         ps.close();
                 } catch (Throwable t) {
                         t.printStackTrace();
                 }
         }

        return handles ;
    }
}
```

APPENDIX 4. TWEET METRIC COLLECTION APPLICATION

```
package org.olumide;
/* application for collecting stats from twitter */

import java.io.*;
import java.util.*;
import java.text.*;
import org.apache.log4j.Logger;
import com.google.gson.*;
import java.sql.Connection;
import java.sql.PreparedStatement;



public class Main{
        private static Logger _log = Logger.getLogger(Main.class.getName());

       private static final Gson gson = new GsonBuilder()
                                .setPrettyPrinting()
                              .setFieldNamingPolicy(
                              FieldNamingPolicy.LOWER_CASE_WITH_UNDERSCORES)
                                .create();


       private static final String twurlCmd =
"/1.1/friends/list.json?cursor=1&screen_name=AdebayoOlumide&skip_status=true&include_user_entitie
s=false&count=200" ;


       private static Connection connection = null;

       public static void main(String[] args) throws Exception{

               Runtime rt = Runtime.getRuntime();
               String _command =  twurlCmd;

               connection = getDB();

               _log.info( _command);

               Process proc = rt.exec(new String[]{"/usr/local/bin/twurl",_command});

               BufferedReader stdInput =
                       new BufferedReader(new InputStreamReader(proc.getInputStream()));
               BufferedReader stdError =
                       new BufferedReader(new InputStreamReader(proc.getErrorStream()));

               // read the output from the command
               _log.info("Here is the standard output of the command:\n");
               String s = null;

               StringBuffer tweets = new StringBuffer();

               while ((s = stdInput.readLine()) != null) {
                       //_log.info(s);
                       tweets.append(s);
               }
```

```
        try{

                JsonArray jarray = parse(tweets.toString());
                if( jarray != null && jarray.size()>=1){
                        for(int i=0;i<jarray.size();i++){
                                process(jarray.get(i),i);
                        }
                }else{
                        _log.debug(" no data A");
                }
        }catch(Exception e){
                e.printStackTrace();
        }

        // read any errors from the attempted command
        _log.info("Here is the standard error of the command (if any):\n");
        while ((s = stdError.readLine()) != null) {
                _log.error(s);
        }

        closeDB(connection);
}

private static void process(JsonElement je,int idx){
        if( je == null){
                return;
        }

        _log.debug(je);

        JsonObject jobj = je.getAsJsonObject();

        String handle = jobj.get("screen_name").getAsString();
        String twitter_id= jobj.get("id_str").getAsString();
        String created = jobj.get("created_at").getAsString();
        created = getDate(created);

        int favorite_count = 0;
        try{ favorite_count= jobj.get("favourites_count").getAsInt(); }catch(Exception e){
}
        int follower_count = 0;
        try{ follower_count= jobj.get("followers_count").getAsInt(); }catch(Exception e){ }
        int friend_count = 0;
        try{ friend_count= jobj.get("friends_count").getAsInt(); }catch(Exception e){ }
        int status_count = 0;
        try{ status_count= jobj.get("statuses_count").getAsInt(); }catch(Exception e){ }
        int listed_count = 0;
        try{ listed_count= jobj.get("listed_count").getAsInt(); }catch(Exception e){ }

        String sql = "insert ignore into twitter_stats ";
        sql += "(handle,created ,twitter_id ,favorite_count";
        sql += " ,follower_count,friend_count,status_count,listed_count,collection_date) ";
        sql += " values ";
        sql += "( ?,?,?,?,?,?,?,?,date(now())) ";
```

```
        String delim=" _";
        StringBuffer _buf = new StringBuffer();
        _buf.append(handle);
        _buf.append(delim);
        _buf.append(created);
        _buf.append(delim);
        _buf.append(twitter_id);
        _buf.append(delim);
        _buf.append(favorite_count);
        _buf.append(delim);
        _buf.append(follower_count);
        _buf.append(delim);
        _buf.append(friend_count);
        _buf.append(delim);
        _buf.append(status_count);
        _buf.append(delim);
        _buf.append(listed_count);

        _log.info(sql);
        _log.info(_buf.toString());

        PreparedStatement ps = null;
         try {
                 ps = connection.prepareStatement(sql);

                ps.setString(1, handle);
                ps.setString(2, created);
                ps.setString(3,twitter_id);
                ps.setLong(4,favorite_count);
                ps.setLong(5,follower_count);
                ps.setLong(6,friend_count);
                ps.setLong(7,status_count);
                ps.setLong(8,listed_count);

                ps.execute();
         } catch (Throwable t) {
                 t.printStackTrace();
         } finally {
                 try {
                         ps.close();
                 } catch (Throwable t) {
                         t.printStackTrace();
                 }
         }
}

private static JsonArray parse(String s) throws Exception{
        JsonElement elem = new JsonParser().parse(s);
        JsonObject obj = elem.getAsJsonObject();

        return obj.getAsJsonArray("users");
}

private static final String TWITTER_DATE="EEE MMM dd HH:mm:ss ZZZZZ yyyy";

private static SimpleDateFormat sf = new SimpleDateFormat(TWITTER_DATE,Locale.ENGLISH);
```

```
        private static String getDate(String s){
                try{

                sf.setLenient(true);
                Date date = sf.parse(s);
                String newstring = new SimpleDateFormat("yyyy-MM-dd").format(date);
                return newstring;
                }catch(Exception e){
                        e.printStackTrace();
                        return s;
                }

        }



        private static void closeDB(Connection conn) throws Exception{
                conn.close();
        }
        private static Connection getDB() throws Exception{
                String url = "jdbc:mysql://localhost:3306/<DATABASE NAME>";
                String username = "<USER>";
                String password = "<PASSWORD>";

                try {

                        _log.info("loading driver!");
                        Class.forName("com.mysql.jdbc.Driver");

                        _log.info("Connecting database...");
                        Connection connection = java.sql.DriverManager.getConnection(
                                url, username, null) ;

                        _log.info("Database connected!");
                        return connection;

                } catch (Exception e) {
                        throw new IllegalStateException("Cannot connect the database!", e);
                }

        }


}
```

APPENDIX 5. AUTOML CLASSIFICATION APPLICATION

```
package org.olumide;

/*
Main class for driving autumn classification process
*/

import java.io.*;
import java.util.*;
import java.text.*;
import org.apache.log4j.Logger;


public class BetaLabellerDriver {
        private static Logger _log = Logger.getLogger(LabellerDriver.class.getName());

        private static final String computeRegion="us-central1";
        private static final String projectId="[PROJECTID";
        private static final String datasetName="[DATASETNAME]";
        private static final String dataSetPath="[DATASETPATH]";
        private static final String modelName = "[MODELNAME]";
        private static final String modelId= "[MODELID]";

        public static void main(String[] args) throws Exception{

                BetaMysqlTweetSource mts = new BetaMysqlTweetSource();
                HashMap<String,String> labels = new HashMap<>();
                HashMap<String,String> tweets = mts.getTweets();


                if( tweets != null && tweets.size()>0){

                        for (Map.Entry<String, String> entry : tweets.entrySet()) {
                                String key = entry.getKey();
                                String value = entry.getValue();


                                String label = predictCategory( modelId,value ) ;
                                _log.debug(key+" :"+value+ " --> "+label);
                                mts.categorize(key,label);
                        }
                }
        }


        private static String predictCategory(String modelId,String tweet){

                try{
                        return PredictionDriver.predict( projectId, computeRegion, modelId, tweet)
;
                }catch(Exception e){
                        e.printStackTrace();
                }

                return null;
        }

}
```

```
package org.olumide;

import java.io.*;
import java.util.*;
import java.text.*;
import org.apache.log4j.Logger;
import com.google.gson.*;
import java.sql.Connection;
import java.sql.PreparedStatement;


public class BetaMysqlTweetSource {
        private static Logger _log = Logger.getLogger(BetaMysqlTweetSource.class.getName());

        private Connection connection = null;

        public HashMap<String,String> getTweets() throws Exception{
                connection = getDB();
                HashMap<String,String> tweets =  fetchTweets();
                closeDB(connection);
                return tweets;
        }

        private void closeDB(Connection conn) throws Exception{
                conn.close();
        }
        private Connection getDB() throws Exception{
                String url = "jdbc:mysql://localhost:3306/[DATABASE]";
                String username = "[USER]";
                String password = "[PASSWORD]";

                try {
                        Class.forName("com.mysql.jdbc.Driver");
                        Connection connection=java.sql.DriverManager.getConnection(
                                url, username, null) ;
                        return connection;
                } catch (Exception e) {
                        throw new IllegalStateException("Cannot connect the database!", e);
                }
        }

        private HashMap<String,String> fetchTweets(){

                HashMap<String,String> tweets= new HashMap<>();

                String sql = "select id,tweet from furniture_tweets where ";
                sql += " id not in (select tweetid from
                        better_furniture_tweets_category)”;
                sql += “ and created>='2019-01-01' limit 21000 ";

                String response = null;
                java.sql.ResultSet rs = null;
                PreparedStatement ps = null;
                try {
                         ps = connection.prepareStatement(sql);
                        rs = ps.executeQuery();
                         while(rs.next()) {
                                 tweets.put(rs.getString("id"), rs.getString("tweet"));
                         }
                 } catch (Throwable t) {
                         t.printStackTrace();
                 } finally {
                         try { rs.close(); } catch (Throwable t) { t.printStackTrace(); }
                         try { ps.close(); } catch (Throwable t) { t.printStackTrace(); }
                 }

                return tweets;
        }

        public void categorize(String tweetid, String cat ) throws Exception{
                connection = getDB();
```

```
            String sql =
                   " insert into better_furniture_tweets_category(tweetid,cat) values(?,?) ";

            PreparedStatement ps = null;
            try{
                     ps = connection.prepareStatement(sql);

                     ps.setString(1, tweetid);
                     ps.setString(2, cat);
                     ps.execute();
             } catch (Throwable t) {
                     t.printStackTrace();
             } finally {
                     try {
                             ps.close();
                     } catch (Throwable t) {
                             t.printStackTrace();
                     }
             }
            closeDB(connection);
        }
}
```

```java
package org.olumide;

// Imports the Google Cloud client library
import com.google.cloud.automl.v1beta1.AnnotationPayload;
import com.google.cloud.automl.v1beta1.ExamplePayload;
import com.google.cloud.automl.v1beta1.ModelName;
import com.google.cloud.automl.v1beta1.PredictResponse;
import com.google.cloud.automl.v1beta1.PredictionServiceClient;
import com.google.cloud.automl.v1beta1.TextSnippet;

import java.io.IOException;
import java.io.PrintStream;

import java.nio.file.Files;
import java.nio.file.Paths;
import java.util.HashMap;
import java.util.Map;

import net.sourceforge.argparse4j.ArgumentParsers;
import net.sourceforge.argparse4j.inf.ArgumentParser;
import net.sourceforge.argparse4j.inf.ArgumentParserException;
import net.sourceforge.argparse4j.inf.Namespace;
import net.sourceforge.argparse4j.inf.Subparser;
import net.sourceforge.argparse4j.inf.Subparsers;


public class PredictionDriver {

        public static String predict( String projectId, String computeRegion, String modelId, String
tweet) throws IOException {

                Float pScore = Float.MIN_VALUE;
                String pName = "nan";

                // Create client for prediction service.
                try (PredictionServiceClient predictionClient = PredictionServiceClient.create()) {

                        ModelName name = ModelName.of(projectId, computeRegion, modelId);

                        // Read the file content for prediction.
                        String content = tweet;

                        // Set the payload by giving the content and type of the file.
                        TextSnippet textSnippet =
                                TextSnippet.newBuilder().setContent(content)
                                .setMimeType("text/plain").build();
                        ExamplePayload payload = ExamplePayload.newBuilder()
                                .setTextSnippet(textSnippet).build();

                        Map<String, String> params = new HashMap<String, String>();
                        PredictResponse response = predictionClient.predict(name, payload, params);

                        for (AnnotationPayload annotationPayload : response.getPayloadList()) {
                                String label = annotationPayload.getDisplayName();
                                Float value = annotationPayload.getClassification().getScore();

                                if( value.compareTo( pScore ) > 0){
                                        pScore = value;
                                        pName = label;
                                }
                        }
                }
        return pName;
        }
}
```

APPENDIX 6. . TWEET SENTIMENT ANALYSIS PYTHON APPLICATION

```
#!/usr/bin/env python3
# coding: utf-8

import nltk
import pandas as pd
import logging
import traceback
import json
import csv


from textblob import TextBlob

logging.basicConfig(level=logging.DEBUG)
logger = logging.getLogger(__name__)

raw = pd.read_csv('tweets.csv',header=None)

categoryDict={}
dmuDict={}
dmucat={}

textIndex=3
for x in range(raw.shape[0]):

        handle=(raw.loc[x].values[0])
        tweetid=(raw.loc[x].values[1])
        text=(raw.loc[x].values[textIndex])
        category=(raw.loc[x].values[4])
        dmu=(raw.loc[x].values[5])

        if category in categoryDict.keys():
                k=1
        else:
                categoryDict[category]={}

        if dmu in dmuDict.keys():
                k=1
        else:
                dmuDict[dmu]={}

        if dmu in dmucat.keys():
                k=1
        else:
                dmucat[dmu]={}

        if category in dmucat[dmu].keys():
                k = 1
        else:
                dmucat[dmu][category]={}


        try:
                blob1 = TextBlob(raw.loc[x].values[textIndex])
                polar=blob1.sentiment.polarity


                if polar > 0:

                        if "positive" in dmuDict[dmu]:
                                dmuDict[dmu]["positive"] += 1;
                        else:
                                dmuDict[dmu]["positive"] = 1;
```

```
                if "positive" in categoryDict[category]:
                        categoryDict[category]["positive"] += 1;
                else:
                        categoryDict[category]["positive"] = 1;

                if "positive" in dmucat[dmu][category]:
                        dmucat[dmu][category]["positive"] += 1;
                else:
                        dmucat[dmu][category]["positive"] = 1;

        elif polar == 0:
                if "neutral" in dmuDict[dmu]:
                        dmuDict[dmu]["neutral"] += 1;
                else:
                        dmuDict[dmu]["neutral"] = 1;

                if "neutral" in categoryDict[category]:
                        categoryDict[category]["neutral"] += 1;
                else:
                        categoryDict[category]["neutral"] = 1;

                if "neutral" in dmucat[dmu][category]:
                        dmucat[dmu][category]["neutral"] += 1;
                else:
                        dmucat[dmu][category]["neutral"] = 1;

        else:
                if "negative" in dmuDict[dmu]:
                        dmuDict[dmu]["negative"] += 1;
                else:
                        dmuDict[dmu]["negative"] = 1;

                if "negative" in categoryDict[category]:
                        categoryDict[category]["negative"] += 1;
                else:
                        categoryDict[category]["negative"] = 1;

                if "negative" in dmucat[dmu][category]:
                        dmucat[dmu][category]["negative"] += 1;
                else:
                        dmucat[dmu][category]["negative"] = 1;

    except Exception as error:
        print("");

        print(x)
        print("some error occured")
        logger.exception(error)

        break

def print_simple_csv(dict,filename):

    f = csv.writer(open(filename, "w"))

    # Write CSV Header
    f.writerow(["anchor", "positive", "negative", "neutral" ])

    for anchor in dict:
        x=dict[anchor]
        pos=0
        neg=0
        neut=0

        if "positive" in x:
            pos=x["positive"]
        if "negative" in x:
            neg=x["negative"]
        if "neutral" in x:
            neut=x["neutral"]
```

```
                f.writerow([anchor, pos, neg,neut ])


def print_dmucat_csv(dict):

        f = csv.writer(open("dmucatdict.csv", "w"))

        # Write CSV Header
        f.writerow(["dmu","cat", "positive", "negative", "neutral" ])

        for dmu in dict:
                for cat in dict[dmu]:

                        x=dict[dmu][cat]
                        pos=0
                        neg=0
                        neut=0

                        if "positive" in x:
                                pos=x["positive"]
                        if "negative" in x:
                                neg=x["negative"]
                        if "neutral" in x:
                                neut=x["neutral"]

                        f.writerow([dmu,cat, pos, neg,neut ])


print_simple_csv(dmuDict,"dmu-dict.csv")
print_simple_csv(categoryDict,"category-dict.csv")
print_dmucat_csv(dmucat)
```